RESEARCH

Matthias Rolffs

Kausalität und mentale Verursachung

Eine Verteidigung des nicht-reduktiven Physikalismus

ABHANDLUNGEN ZUR **PHILOSOPHIE**

OPEN ACCESS



J.B. METZLER

Abhandlungen zur Philosophie

In dieser Reihe erscheinen Monographien und Sammelbände zur Philosophie bzw. zu angrenzenden oder die Fachgrenze überschreitenden Themen. Klassische Gebiete sollen neu abgesteckt, aktuelle Felder bearbeitet und innovative Fragen formuliert und zur Diskussion gestellt werden.
Wir freuen uns über Ihr Interesse und Ihren Vorschlag!

Matthias Rolffs

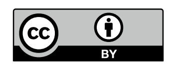
Kausalität und mentale Verursachung

Eine Verteidigung des nicht-reduktiven Physikalismus



Matthias Rolffs
Bern, Schweiz

Publiziert mit Unterstützung des Schweizerischen Nationalfonds zur Förderung der wissenschaftlichen Forschung.
Die Druckvorstufe dieser Publikation wurde vom Schweizerischen Nationalfonds zur Förderung der wissenschaftlichen Forschung unterstützt.



ISSN 2945-9923 ISSN 2945-9931 (electronic)
Abhandlungen zur Philosophie
ISBN 978-3-662-66777-4 ISBN 978-3-662-66778-1 (eBook)
https://doi.org/10.1007/978-3-662-66778-1

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

Planung/Lektorat: Marija Kojic
J.B. Metzler ist ein Imprint der eingetragenen Gesellschaft Springer-Verlag GmbH, DE und ist ein Teil von Springer Nature.
Die Anschrift der Gesellschaft ist: Heidelberger Platz 3, 14197 Berlin, Germany

Danksagung

Diese Arbeit basiert auf meiner Dissertation, die ich am 27.08.2020 an der Universität Bonn verteidigt habe.

Besonderer Dank gilt meinem Erstbetreuer Andreas Bartels, der mich bereits seit meinem Bachelorstudium unterstützt und gefördert hat. Durch seine Kommentare und die zahlreichen Diskussionen mit ihm habe ich viele wertvolle Einsichten gewonnen, ohne die dieses Buch nicht wäre, was es ist. Andreas Hüttemann danke ich für die Zweitbetreuung meiner Dissertation sowie für sein Gutachten, das hilfreiche Anmerkungen zu einer früheren Fassung dieses Textes enthält.

Auf dem Weg von der ersten Idee zur Veröffentlichung haben mich zahlreiche weitere Menschen unterstützt, beeinflusst oder inspiriert. Stefan Heidl, Vera Hoffmann-Kolss, Elisabeth Muchka, Kian Salimkhani, Alexander Samans und Jonas Werner haben Auszüge des Manuskripts gelesen und mit hilfreichen Anmerkungen versehen. Angelika Mann und Elisabeth Rolffs haben die gesamte Arbeit korrekturgelesen und mich bei der Erstellung des Literaturverzeichnisses unterstützt. Auch hatte ich mehrfach die Gelegenheit, Teile der Arbeit in Forschungskolloquien in Bonn, Köln, Bologna und Bochum sowie bei verschiedenen Konferenzvorträgen zur Diskussion zu stellen. Hierbei habe ich wertvolle Inputs u. a. von Andreas Bartels, Boris Brandhoff, Raffaela Campaner, Sabrina Coninx, Leonard Dung, Jochen Faseler, Cord Friebe, Stefan Heidl, Vera Hoffmann-Kolss, Andreas Hüttemann, Dennis Lehmkuhl und Kian Salimkhani erhalten. Schließlich habe ich von Diskussionen mit Teilnehmer*innen von themenverwandten Seminaren profitiert, die ich an den Universitäten Bonn und Bern gegeben habe. All diesen Personen möchte ich herzlich danken!

Dankbar bin ich des Weiteren für die finanzielle Unterstützung durch das Cusanuswerk. Auch die Graduierten- und Fachschaftstagungen während meiner Zeit in der Promotionsförderung waren sehr bereichernd und haben diese Arbeit mitgeprägt. Des Weiteren danke ich dem Schweizer Nationalfonds für die finanzielle Unterstützung der Open Access Publikation.

Schließlich bin ich meinen Freund*innen und meiner Familie für die emotionale Unterstützung während der gesamten Promotionszeit dankbar. Besonderer Dank gilt dabei Elisabeth Rolffs, Detlef Rolffs, Angelika Mann, Lina Rolffs, Jan Baumann, Felix und Fabian Rolffs, Rainer Rolffs, Jasmin Rolffs, Clara, Anton und Nico Rolffs, Kian Salimkhani und Alexander Samans.

Inhaltsverzeichnis



Einleitung 1

In dieser Arbeit verfolge ich das Ziel, eine weit verbreitete Position in der Philosophie des Geistes gegen einen einflussreichen Einwand zu verteidigen. Die Position, die ich verteidigen möchte, ist der nicht-reduktive Physikalismus. Der Einwand, gegen den ich diese Position verteidigen möchte, ist der Vorwurf des Epiphänomenalismus.

Der nicht-reduktive Physikalismus besagt im Kern, dass mentale – und, wie wir sehen werden, auch andere ‚höherstufige' – Eigenschaften zwar nicht mit physischen Eigenschaften identisch sind, sie aber dennoch so eng an physische Eigenschaften gebunden sind, dass man sinnvollerweise von einer auf grundlegender Ebene rein physischen Welt sprechen kann. Die genauere Interpretation dieser Grundidee wird Teil 2 dieser Arbeit ausmachen.

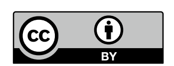
Der Vorwurf des Epiphänomenalismus wendet gegen den nicht-reduktiven Physikalismus ein, dass aus seiner Grundidee folgt, dass es keine mentale Verursachung gibt: Wenn mentale Eigenschaften tatsächlich nicht mit physischen Eigenschaften identisch sind, sondern lediglich eng an sie gebunden sind, dann können mentale Eigenschaften niemals kausal wirksam werden. Es ist dann nicht mehr möglich, zu behaupten, dass Überzeugungen und Wünsche Verhalten verursachen. Diese Konsequenz könne nicht in Kauf genommen werden und der nicht-reduktive Physikalismus müsse daher aufgegeben werden. Die Formulierung und Diskussion dieses Einwands wird Teil 3 dieser Arbeit ausmachen.

Meine Erwiderung auf den Vorwurf des Epiphänomenalismus schließlich beruht auf einer eingängigen Beschäftigung mit der Metaphysik der Kausalität. Meines Erachtens liegt der Schlüssel zur Zurückweisung des Vorwurfs des Epiphänomenalismus in einer Unterscheidung zwischen verschiedenen Begriffen der Kausalität. Tatsächlich ist der nicht-reduktive Physikalismus darauf festgelegt, dass es in einem bestimmten Sinne keine mentale Verursachung gibt:

M. Rolffs, *Kausalität und mentale Verursachung*, Abhandlungen zur Philosophie,
https://doi.org/10.1007/978-3-662-66778-1_1

Mentale Ereignisse *bringen* ihre Verhaltenswirkungen nicht im selben Sinne *hervor*, wie vielleicht die Bewegung einer Billardkugel die Bewegung einer anderen Billardkugel hervorbringt. Hieraus, so meine ich, folgt jedoch nichts, was zu einer Aufgabe des nicht-reduktiven Physikalismus zwingt. Was hingegen zu einer Aufgabe des nicht-reduktiven zwingen würde, wäre die These, dass mentale Ereignisse *keinen Unterschied* für den Verlauf der Welt machen. Wenn aus dem nicht-reduktiven Physikalismus folgen würde, dass alle Menschen sich genauso verhalten würden, wie sie sich tatsächlich verhalten, selbst wenn sie gar keine Gedanken, Überzeugungen und Wünsche hätten, müsste die Position natürlich aufgegeben werden. Diese inakzeptable Konsequenz hat der nicht-reduktive Physikalismus, wie ich zeigen werde, jedoch keinesfalls. Diese Erwiderung arbeite ich in Teil 4 dieser Arbeit detailliert aus.

Obgleich die drei Teile dieser Arbeit aufeinander aufbauen und die ersten beiden Teile zu einem guten Stück eine begriffsklärende und vorbereitende Funktion haben, stellt jeder Teil auch einen Beitrag zu einer laufenden Debatte dar und kann als solcher eigenständig gelesen werden. Teil 2 zeigt auf, dass der nicht-reduktive Physikalismus eine konsistente und attraktive Position ist, die klar von anderen Positionen abgegrenzt werden kann. Schon dies wird jedoch von einigen Autor*innen angezweifelt. Teil 3 enthält einige interessante Punkte zur Formulierung des Vorwurfs des Epiphänomenalismus und unterscheidet eine Reihe von Exklusionsargumenten, die in der Literatur nicht sauber voneinander unterschieden werden. Teil 4 schließlich entwickelt detailliert eine Erwiderung auf den Vorwurf des Epiphänomenalismus, die auf einer Unterscheidung zwischen zwei Begriffen der Kausalität beruht.





2 Der nicht-reduktive Physikalismus

Die Grundidee des Physikalismus besteht darin, dass physischen Entitäten ein ontologischer Sonderstatus zukommt: Alles, was es gibt, ist letztlich physisch. Der *nicht-reduktive* Physikalismus ist dadurch ausgezeichnet, dass er diese Grundidee auf eine charakteristische Weise expliziert. Insbesondere lässt der nicht-reduktive Physikalismus im Gegensatz zu anderen Versionen des Physikalismus Spielraum für die Autonomie der nicht-physikalischen Spezialwissenschaften und der (Alltags-)Psychologie.

Der nicht-reduktive Physikalismus ist eine moderate und vermittelnde Position: Er ist durch das Bemühen um einen Zwischenweg zwischen zwei unwillkommenen Alternativen ausgezeichnet: Einerseits haben gewisse Entwicklungen in der Philosophie des Geistes in der zweiten Hälfte des 20. Jahrhunderts dazu geführt, dass *reduktive* Versionen des Physikalismus in Verruf geraten sind. Ausschlaggebend sind hier insbesondere Überlegungen zur multiplen Realisierbarkeit mentaler und spezialwissenschaftlicher Eigenschaften.[1] Auf der anderen Seite sollte die Ablehnung des reduktiven Physikalismus nicht zu einem echten *Dualismus* führen, in dem der mentale Bereich der Wirklichkeit strikt von dem physischen Bereich getrennt ist. Denn eine solche strikte Trennung lässt es geheimnisvoll erscheinen, wie mentale Eigenschaften in eine Welt passen, die doch auf fundamentaler Ebene durch physikalische Gesetze bestimmt zu sein scheint. Eine Position, die die Grundidee des Physikalismus rettet, ohne aber auf problematische, reduktionistische Annahmen angewiesen zu sein, scheint vor diesem Hintergrund attraktiv.

[1] Kim (1992: 3) schreibt über die Bedeutung der These der multiplen Realisierung: „MR has not only ushered in ‘nonreductive physicalism’ as the new orthodoxy on the mind-body problem, but in the process has put the very word „reductionism“ in disrepute, making reductionisms of all stripes an easy target of disdain and curt dismissals“.

M. Rolffs, *Kausalität und mentale Verursachung*, Abhandlungen zur Philosophie,
https://doi.org/10.1007/978-3-662-66778-1_2

Die Vertretbarkeit eines solchen Zwischenwegs ist jedoch keine Selbstverständlichkeit. Tatsächlich steht der nicht-reduktive Physikalismus im Verdacht, sich nicht klar von den reduktiven und dualistischen Alternativen abgrenzen zu können. So argumentiert zum Beispiel Jaegwon Kim, dass jede spezifische Formulierung des nicht-reduktiven Physikalismus sich letztlich entweder als eine Version des reduktiven Physikalismus entpuppt, oder in einen echten Dualismus kollabiert.[2]

Ziel dieses Kapitels ist es, im Lichte dieser Problematik eine Formulierung des nicht-reduktiven Physikalismus zu entwickeln, die klar von konkurrierenden Positionen abgegrenzt ist und die als Grundlage für eine Verteidigung des nicht-reduktiven Physikalismus gegen den Vorwurf des Epiphänomenalismus in Teil 3 und Teil 4 dieser Arbeit dienen kann.

Im Folgenden gehe ich hierfür zunächst genauer auf die Grundidee jeder Form des Physikalismus ein und treffe eine vorläufige Unterscheidung zwischen verschiedenen Versionen des Physikalismus. Daraufhin formuliere ich drei nicht-reduktive Kernthesen und eine physikalistische Kernthese des nicht-reduktiven Physikalismus (Abschnitt 2.1.). Zur genaueren Explikation dieser Kernthesen erläutere ich zunächst die beiden darin enthaltenen ontologischen Grundbegriffe ‚Eigenschaft' und ‚Ereignis' (Abschnitt 2.2.) und behandle die Frage, wie der Begriff physischer Eigenschaften bestimmt werden kann (Abschnitt 2.3). Hierauf aufbauend werden die nicht-reduktiven Kernthesen (Abschnitt 2.4) und die physikalistische Kernthese (Abschnitt 2.5) diskutiert. Dies ermöglicht eine Abgrenzung des nicht-reduktiven Physikalismus von anderen Versionen des Physikalismus und vom Dualismus (Abschnitt 2.6).

2.1 Physikalismus und nicht-reduktiver Physikalismus

Dieser Abschnitt dient einer ersten groben Einordnung des nicht-reduktiven Physikalismus in einen weiteren Kontext. Hierfür gehe ich in Abschnitt 2.1.1. zunächst knapp auf die Art von Fragen ein, auf die der Physikalismus im Allgemeinen sowie der nicht-reduktive Physikalismus im Speziellen eine Antwort geben sollen. Daraufhin skizziere ich in Abschnitt 2.1.2. die Grundidee des Physikalismus als einer globalen ontologischen These. In Abschnitt 2.1.3. schildere

[2] Vgl. z. B. Kim (1989: 40): „[I]f a relation is weak enough to be nonreductive, it tends to be too weak to serve as a dependence relation; conversely, when a relation is strong enough to give us dependence, it tends to be too strong – strong enough to imply reducibility".

ich, wie diese globale ontologische These in ihrer Reichweite eingeschränkt werden kann. Dies bildet die Grundlage für eine erste grobe Charakterisierung des nicht-reduktiven Physikalismus in Abschnitt 2.1.4.

2.1.1 Metaphysische Fragen

In der Metaphysik werden grob gesprochen folgende Fragen behandelt:[3]

Existenz: Was gibt es?
Beschaffenheit: Wie sind die Dinge, die es gibt, beschaffen?
Beziehung: Wie hängen die Dinge, die es gibt, miteinander zusammen?
Modalität: Was ist möglich und notwendig?

Man könnte zunächst versuchen, die *Existenz-Frage* im Stile einer Auflistung zu beantworten: Es gibt (zum Beispiel) Zahlen, mathematische Formeln, alltagssprachliche Sätze mit Bedeutungen und Wahrheitswerten, Propositionen, vage Andeutungen, Gegenstände, an denen man sich die Zehen stoßen kann, wie Tische, Stühle oder Türrahmen, Zehen, Arme und Beine, Gehirne, Gefühle, Überzeugungen und Wünsche, Geldscheine, Finanzkrisen und Geschlechterrollen, Farben und ihre Schattierungen, moralische Wahrheiten, Möglichkeiten und Wahrscheinlichkeiten, Personen, freie und unfreie Handlungen, die unter den gegeben Umständen vernünftigste Handlungsoption, Wolken, Warenwerte, Gene, Raum und Zeit, Kausalität, Naturgesetze, Elektronen, Positronen und Neutronen, schwache und starke Wechselwirkungen, Gravitation usw. Auf diese Weise könnte man einige tausend Seiten füllen und würde doch nichts Erhellendes (geschweige denn Erschöpfendes) sagen.[4] Die Aufgabe ontologischer Theorien ist es daher nicht, lange Listen zu schreiben, sondern ein wenig Ordnung in die Vielfalt zu bringen.

Eine Strategie, Ordnung in eine solche Liste zu bringen, besteht in der Behandlung der *Beschaffenheits-Frage*: Indem man spezifiziert, *wie* die Einträge der Liste beschaffen sind und *als was* sie existieren, lassen sich Gemeinsamkeiten

[3] Vgl. für eine ähnliche Charakterisierung der Fragestellungen der Metaphysik: Bartels (Manuskript).

[4] Man würde sicherlich nicht umhinkommen, auf dem Weg einige *kontroverse* Dinge zu sagen. So würde man sicherlich schon auf der Straße einige Leute finden, die bestreiten, dass es moralische Wahrheiten gibt. Ebenso findet man bei genauer Recherche für jeden der oben aufgeführten Einträge in unsere Liste sicherlich einen Philosophen, der diesen Eintrag gerne streichen würde.

zwischen verschiedenen Einträgen erkennen. So kann man die verschiedenen Einträge in Gruppen oder Kategorien einteilen: Es gibt Eigenschaften (wie Farben), Ereignisse (wie Finanzkrisen) und Gegenstände (wie Tische); es gibt abstrakte Entitäten (wie Zahlen) und konkrete Entitäten (wie Tische); es gibt mentale Entitäten (wie Überzeugungen und Wünsche) und physische Entitäten (wie Elektronen); usw.

Eine zweite Strategie zur Schaffung von Ordnung ergibt sich aus der *Beziehungs-Frage*: So kann die These vertreten werden, dass nicht alle Entitäten gleichermaßen eigenständig existieren und manche Entitäten *ontologisch abhängig* sind von anderen Entitäten. Mit einer solchen Abhängigkeitsbeziehung einher geht die Idee, dass sich eine relativ kleine Menge von *basalen* oder *fundamentalen* Entitäten identifizieren lässt, von denen alle weiteren Entitäten abhängig sind. In diesem Sinne schreibt etwa Frank Jackson:

> Metaphysics is concerned with what there is and what it is like. But it is not concerned with any old shopping list of what there is and what it is like. Metaphysicians seek a comprehensive account of some subject matter – the mind, the semantic, or, most ambitiously, everything – in terms of a limited number of more or less basic notions.[5]

Solche Abhängigkeitsbeziehungen und die damit einhergehende Idee der Fundamentalität führen zugleich zu einer Auseinandersetzung mit der *Modalitäts-Frage*: Denn wenn etwa eine bestimmte Anordnung von Elementarteilchen als metaphysische Basis für die Existenz eines Tisches dient, dann *erzwingt* die Existenz dieser Anordnung von Elementarteilchen wohl zugleich die Existenz des Tisches: Es *kann* gar nicht sein, dass die Elementarteilchen ‚tischmäßig' angeordnet sind, es aber keinen Tisch gibt.

2.1.2 Globaler Physikalismus: Grundidee

Der Physikalismus ist eine ontologische Theorie in diesem Sinne: Er ist eine Theorie darüber, was es alles gibt, wie alles beschaffen ist und wie alles miteinander zusammenhängt. Die Grundthese des Physikalismus kann in einer ersten Annäherung wie folgt formuliert werden:

---

[5] Jackson (1994: 25), Jackson (1998: 4).

Globaler Physikalismus: Alles, was es gibt, ist *letztlich* physisch.

Der Physikalismus stellt also eine Bedingung an alle Entitäten, die in eine Ontologie aufgenommen werden: Sie sollen ‚*letztlich* physisch' sein.[6]

Die obige Liste enthält einige Einträge, die *offenbar* physisch sind. Insbesondere die letzten Einträge, die sich direkt gegenwärtigen physikalischen Theorien verdanken (Elektronen, Positronen und Neutronen, schwache und starke Wechselwirkungen sowie Gravitation), sind in einem klaren und nicht abgeleiteten Sinne physisch. Dasselbe könnte man vielleicht von den Gegenständen sagen, an denen man sich die Zehen stoßen kann.

Die Liste enthält aber auch eine ganze Reihe von Einträgen, deren physische Natur keinesfalls offenbar ist (Zahlen, Finanzkrisen, Überzeugungen, moralische Wahrheiten, Bedeutungen, Wahrheitswerte usw.). Die grundsätzliche Herausforderung für den Physikalismus besteht im Umgang mit diesen anderen Einträgen.[7]

Hierbei stehen Physikalist*innen drei Herangehensweisen offen: Erstens können sie einen Eintrag, der nicht offenbar physisch ist, schlicht von der Liste streichen. Das problematische Phänomen ist dann zwar nicht physisch, existiert aber auch gar nicht und stellt daher kein Problem für die physikalistische Grundthese dar. Dies kann der *Weg der Elimination* genannt werden. Zweitens können Physikalist*innen ein problematisches Phänomen mit einem physischen Phänomen identifizieren. Entgegen dem ersten Anschein ist das problematische Phänomen dann doch selbst physisch. Dies kann der *Weg der Identifikation* genannt werden. Drittens schließlich können sie aufzeigen, dass das problematische Phänomen sich unweigerlich aus physischen Phänomenen ergibt, so dass es sich nicht um ein ontologisch eigenständiges, über die physischen Phänomene hinausgehendes Phänomen handelt. In diesem Falle ist das problematische Phänomen zwar selbst nicht im eigentlichen Sinne physisch, wohl aber in einem hinreichend starken Sinne von physischen Phänomenen ontologisch abhängig

[6] Vgl. auch Stoljar (2015).

[7] Ein uneingeschränkt globaler Physikalismus, der auch Entitäten wie Zahlen betrifft, wird nicht häufig vertreten. Vgl. hierzu auch Nimtz (2009: 2) und Stoljar (2010: 30). Welche Einschränkungen für ‚den' Physikalismus ausschlaggebend sind, ist nicht eindeutig. Im Prinzip gibt es ebenso viele Versionen des Physikalismus wie mögliche Einschränkungen. In jedem Falle ist es daher besser, in einer Untersuchung des Physikalismus die intendierten Einschränkungen explizit zu machen – was ich im folgenden Abschnitt tue.

und daher doch *letztlich* physisch.[8] Dies kann der *Weg der Abhängigkeit* genannt werden.

In einer ersten groben Einteilung können verschiedene Versionen des Physikalismus danach unterschieden werden, welche Herangehensweise für den Umgang mit problematischen Phänomenen sie wählen: Der *eliminative Physikalismus* bevorzugt den Weg der Elimination, der *reduktive Physikalismus* den Weg der Identifikation und der *nicht-reduktive Physikalismus* den Weg der Abhängigkeit.[9] Im Mittelpunkt dieser Arbeit steht der nicht-reduktive Physikalismus.

2.1.3 Einschränkungen des Physikalismus

Die Version des Physikalismus, die in dieser Arbeit thematisiert wird, lässt sich noch weiter spezifizieren. Denn bis hierher habe ich den Physikalismus als eine *uneingeschränkt allgemeine* ontologische These spezifiziert. Es gibt jedoch wenigstens drei Arten, wie diese globale These eingeschränkt werden kann.

Erstens kann die These auf Entitäten einer bestimmten ontologischen Kategorie eingeschränkt werden: Sie betrifft dann nicht mehr uneingeschränkt alles, was es gibt, sondern zum Beispiel nur Objekte, nur Ereignisse, oder nur Eigenschaften. So ergibt sich ein *Objekt-Physikalismus*, ein *Ereignis-Physikalismus* oder ein *Eigenschafts-Physikalismus.* Die spätere Diskussion in diesem Kapitel wird sich auf den Eigenschafts-Physikalismus konzentrieren, aus dem nach dem in Abschnitt 2.2 eingeführten System von Kategorien ein Ereignis-Physikalismus folgt.

Zweitens kann die These auf Entitäten eines bestimmten (kategorienübergreifenden) Bereichs eingeschränkt werden. Hiermit ist zum Beispiel eine Einschränkung auf mentale, spezialwissenschaftliche, semantische, moralische oder mathematische Entitäten gemeint. Es ergibt sich so ein *mentaler Physikalismus*, ein *spezialwissenschaftlicher Physikalismus*, ein *semantischer Physikalismus* usw. Die Vielfältigkeit der *lokalen* Physikalismus-Thesen verdeutlicht, wie ehrgeizig die These des *globalen* Physikalismus ist. Eine umfassende Verteidigung des globalen Physikalismus müsste jede einzelne der lokalen Thesen plausibilisieren. Dabei eröffnet jede der lokalen Thesen eine ganz eigene Baustelle, der in der Regel eine eigene Disziplin der Philosophie zugeordnet ist. Tatsächlich

---

[8] Auf die schwierige Frage, wie Eigenschaften in diesem Sinne letztlich physisch sein können, ohne mit physischen Eigenschaften identisch zu sein, gehe ich in Abschnitt 2.5. ausführlich ein.

[9] Diese grobe Einteilung wird in Abschnitt 2.6.5. in einigen Aspekten verfeinert, kann aber als erste Orientierung dienen.

werden in den Debatten um den Physikalismus faktisch in aller Regel lokale Physikalismus-Thesen verhandelt. In dieser Arbeit (sowie in der Debatte, in der diese Arbeit verortet ist) stehen der mentale und der spezialwissenschaftliche Physikalismus im Mittelpunkt.

Drittens schließlich kann und sollte der Physikalismus auf aktuale Entitäten beschränkt werden. Physikalist*innen behaupten typischerweise nicht, dass jede mögliche Entität letztlich physisch ist. Vielmehr ist die These auf diejenigen Entitäten beschränkt, die in der aktualen Welt vorkommen. Physikalist*innen akzeptieren typischerweise, dass es – weit entfernte – mögliche Welten gibt, in denen es kartesische Seelen mit mentalen Eigenschaften gibt, die keinerlei physische Eigenschaften haben. Dieses Zugeständnis kommt der Einsicht gleich, dass der kartesische Dualismus zwar falsch, aber nicht unmöglich ist. Entsprechend ist der Physikalismus laut vielen Physikalist*innen zwar wahr, aber kontingent. Aus diesem Grund wird der Physikalismus in aller Regel auf aktuale Entitäten beschränkt.[10]

Gegenstand dieser Arbeit ist zusammenfassend eine *nicht-reduktive* Version des Physikalismus, die sich auf *aktual instantiierte*, *mentale und spezialwissenschaftliche Eigenschaften* beschränkt. Es ist diese Version des Physikalismus, die ich im Folgenden schlicht mit ‚nicht-reduktiver Physikalismus' bezeichnen werde und die in den kommenden Abschnitten weiter erläutert werden soll.

2.1.4 Die Kernthesen des nicht-reduktiven Physikalismus

Im Kern besagt der nicht-reduktive Physikalismus also, dass mentale und spezialwissenschaftliche Eigenschaften nicht mit physischen Eigenschaften identisch sind, sondern ‚lediglich' in einem starken Sinne ontologisch abhängig von physischen Eigenschaften sind. Dieser Kern kann noch etwas weiter konkretisiert werden:

Der nicht-reduktive Physikalismus strebt einen Kompromiss zwischen physikalistischer Ontologie und der Autonomie der Spezialwissenschaften und des Mentalen an. Die Position lässt sich daher gut in ihre nicht-reduktionistischen Kernthesen und in eine physikalistische Kernthese aufteilen. Die nicht-reduktionistischen Kernthesen sollen dabei der Autonomie der Spezialwissenschaften und des Mentalen Rechnung tragen, während die physikalistische Kernthese die Position als eine Version des Physikalismus auszeichnen soll.

---

[10] Vgl. Stoljar (2010: 32–34 und 130–131).

Die nicht-reduktionistischen Kernthesen können wie folgt zusammengefasst werden:

Realismus: Der Realismus in Bezug auf mentale und spezialwissenschaftliche Aussagen und Eigenschaften ist wahr.
Nicht-Identität: Mentale und spezialwissenschaftliche Eigenschaften sind nicht mit physischen Eigenschaften identisch.
Multiple Realisierung: Mentale und spezialwissenschaftliche Eigenschaften sind auf physischer Ebene multipel realisiert.

Diese Thesen teilt der nicht-reduktive Physikalismus mit typischen Versionen anderer nicht-reduktionistischer Positionen wie dem Emergentismus und dem Dualismus. Von diesen Positionen abgegrenzt ist der nicht-reduktive Physikalismus jedoch durch die folgende physikalistische Kernthese:

Ontologische Abhängigkeit: Mentale und spezialwissenschaftliche Eigenschaften sind von physischen Eigenschaften in einem hinreichend starken Sinne ontologisch abhängig.

Die vier Kernthesen des nicht-reduktiven Physikalismus sind hier absichtlich zunächst eher vage formuliert und können jeweils auf unterschiedliche Weisen präzisiert werden. Es ist dabei keinesfalls klar, ob nicht-reduktive Physikalist*innen auf jede mögliche Präzisierung der Thesen festgelegt sind. Die Interpretationsfragen, die mit den jeweiligen Thesen zusammenhängen, sollen in den kommenden Abschnitten diskutiert werden.

Dabei gehe ich zunächst auf zwei allgemeine Fragestellungen ein, die alle fünf Prinzipien betreffen: In Abschnitt 2.2 behandle ich die Frage: Was ist mit den Ausdrücken ‚Eigenschaft' und ‚Ereignis' gemeint? In Abschnitt 2.3 geht es um die Frage: Was hat es mit der Unterscheidung zwischen ‚physischen' und ‚mentalen' bzw. ‚spezialwissenschaftlichen' Eigenschaften auf sich? In Abschnitt 2.4 werden die drei nicht-reduktionistischen Kernthesen des nicht-reduktiven Physikalismus in den Blick genommen. Abschnitt 2.5. schließlich widmet sich der physikalistischen Kernthese des nicht-reduktiven Physikalismus.

2.2 Eigenschaften und Ereignisse

Die für den nicht-reduktiven Physikalismus ausschlaggebenden ontologischen Grundbegriffe sind ‚Eigenschaft' und ‚Ereignis'. Die vier oben formulierten Kernthesen betreffen Eigenschaften. Der Begriff des Ereignisses wird immer dann relevant, wenn es – wie in der These der kausalen Geschlossenheit und im Vorwurf des Epiphänomenalismus – um kausale Beziehungen geht, da Ereignisse üblicherweise als die Relata der Kausalrelation verstanden werden.

Ziel dieses Abschnittes ist es, diese ontologischen Grundbegriffe ansatzweise zu klären, um eine Grundlage für die Diskussion der vier Kernthesen und des Vorwurfs des Epiphänomenalismus zu schaffen. Zunächst gehe ich dabei unter 2.2.1. auf den Eigenschaftsbegriff ein, bevor ich unter 2.2.2. den Ereignisbegriff erläutere.

2.2.1 Eigenschaften

Manche Dinge ähneln einander in der einen oder anderen Hinsicht. Beispielsweise haben mein Laptop und mein Pullover dieselbe Farbe. Sie sind beide schwarz. Andererseits unterscheiden sie sich auch: sie haben eine unterschiedliche Form und Konsistenz. So ist mein Laptop eckig und hart, während mein Pullover weder eckig noch hart ist. Diese simple Beobachtung kann auch dadurch ausgedrückt werden, dass die beiden Objekte bestimmte Eigenschaften teilen und andere Eigenschaften nicht teilen. Das Teilen von Eigenschaften ist demnach eng verbunden mit der Idee, dass sich Objekte in verschiedenen Hinsichten ähneln.

In bestimmten Kontexten wird der Eigenschaftsbegriff auf eine Art und Weise verwendet, die diese Verbindung mit Ähnlichkeitshinsichten nicht beachtet. Es wird dann gesagt, dass jeder beliebigen Menge von Objekten eine Eigenschaft entspricht. Hieraus folgt, dass auch jedes denkbare Prädikat einer Sprache eine Eigenschaft bezeichnet. Nach dieser Auffassung von Eigenschaften gibt es beispielsweise trivialerweise eine Eigenschaft, die nur mein linker Daumen und der Mond teilen, und die einfach darin besteht, mein linker Daumen oder der Mond zu sein. Klarerweise gibt es jedoch keine Ähnlichkeitshinsicht, die meinen linken Daumen mit dem Mond verbindet und diese beiden Objekte von allen anderen Objekten unterscheidet.

Für den Eigenschaftsbegriff, der in einem engen Zusammenhang zu Ähnlichkeitshinsichten steht, hat sich der Ausdruck ‚spärliche Eigenschaft' durchgesetzt. *Spärliche Eigenschaften* sind von so genannten *üppigen Eigenschaften* zu unterscheiden, die im gerade geschilderten Sinne von Ähnlichkeitshinsichten losgelöst

sind.[11] Wenn ich in dieser Arbeit unqualifiziert von Eigenschaften spreche, meine ich damit spärliche Eigenschaften. Wo die Unterscheidung zwischen spärlichen und üppigen Eigenschaften relevant ist, weise ich explizit darauf hin.

Die Verbindung zwischen (spärlichen) Eigenschaften und Ähnlichkeitshinsichten kann wie folgt präzisiert werden:[12]

> **(EA)** Für alle Objekte o, o‘ gilt: o und o‘ teilen genau dann eine Eigenschaft *F*, wenn sich o und o‘ in F-Hinsicht ähneln.

Diese Definition ist nicht zwangsläufig als reduktive Analyse aufzufassen: Es soll hiermit nicht suggeriert werden, dass die Relation der Ähnlichkeit in einer Hinsicht ontologisch primär gegenüber dem Teilen einer Eigenschaft ist. Stattdessen wird mit dieser Definition lediglich die zentrale theoretische Rolle der Rede von Eigenschaften betont: Eigenschaften sind mit Ähnlichkeiten zwischen Objekten verbunden.

Hervorzuheben ist weiterhin, dass die relevante Relation nicht die der Ähnlichkeit *simpliciter* ist, sondern die der Ähnlichkeit *in einer Hinsicht*.[13] Ähnlichkeit simpliciter ist zu grob, um den mit dem Eigenschaftsbegriff verbundenen Intuitionen und Zwecken zufriedenstellend Rechnung zu tragen. Dies wird unter anderem an dem so genannten Problem der nicht-perfekten Gemeinschaft (‚problem of imperfect community‘) deutlich: Eine Menge von Objekten, die sich untereinander *simpliciter* ähneln, muss nicht einer Eigenschaft entsprechen. Dass sich eine Menge von Objekten simpliciter ähnelt, garantiert nicht, dass alle Objekte in der Menge eine Eigenschaft teilen. Beispielsweise ähneln sich ein kleiner, schwarzer Kreis, ein großer, roter Kreis und ein kleines, rotes Quadrat jeweils untereinander. Allerdings teilen diese drei Objekte keine Eigenschaft. Unterscheidet man hingegen Ähnlichkeitshinsichten, lassen sich die Objekte, die eine Eigenschaft teilen, problemlos herausgreifen.[14]

Definition (EA) verlangt nicht, dass sich zwei Objekte, die eine Eigenschaft teilen, in einer Hinsicht *exakt* oder *perfekt* ähneln müssen. Zwei rote Objekte

---

[11] Vgl. Lewis (1983) und Lewis (1986a: 59–61). Bei Lewis ist die Unterscheidung zwischen spärlichen und üppigen Eigenschaften in einer Unterscheidung zwischen natürlichen und nicht-natürlichen Eigenschaften begründet. Zudem findet sich gelegentlich eine Unterscheidung zwischen genuinen und nicht-genuinen Eigenschaften, die ebenfalls dieselbe Funktion hat (vgl. z. B. Loewer (2007)).

[12] Vgl. Audi (2013).

[13] Vgl. für eine Diskussion des Verhältnisses zwischen Ähnlichkeit simpliciter und Ähnlichkeit in einer Hinsicht auch Cowling (2017).

[14] Vgl. Audi (2013: 7).

können sich in ihrer Farbe beispielsweise durchaus unterscheiden, indem sie unterschiedliche rot-Töne haben. Dennoch können wir davon sprechen, dass sie die Eigenschaft teilen, rot zu sein. Auch nicht-perfekte Ähnlichkeiten zwischen Objekten können also die Rede vom Teilen einer Eigenschaft rechtfertigen.[15]

Diese mit (EA) verbundenen Festsetzungen haben Konsequenzen für die Formulierung des nicht-reduktiven Physikalismus, die insbesondere im Zusammenhang mit der Realismus-These und der These der multiplen Realisierung relevant werden. Abgesehen von diesen Festsetzungen hinsichtlich des Eigenschaftsbegriff werde ich im Folgenden neutral gegenüber verschiedenen Theorien über die Metaphysik von Eigenschaften bleiben: Insbesondere ist es nicht nötig, sich auf eine universalienrealistische, tropentheoretische oder streng nominalistische Eigenschaftstheorie festzulegen. Insofern solche Theorien die Rede von Ähnlichkeitshinsichten rekonstruieren können, lassen sich auch die in dieser Arbeit diskutierten Probleme und Positionen formulieren.

2.2.2 Ereignisse

Der zweite grundlegende ontologische Begriff, der für die Formulierung des Physikalismus und des Vorwurfs des Epiphänomenalismus eine zentrale Rolle spielt, ist der Begriff ‚Ereignis'. Ich stelle im Folgenden zwei einflussreiche Explikationen dieses Begriffes vor.

Donald Davidson versteht Ereignisse als „unrepeatable, dated individuals"[16]: Jedes Ereignis tritt genau einmal, zu einer bestimmten Zeit und an einem bestimmten Ort, auf. Ereignisse wiederholen sich nicht und sie sind raumzeitlich lokalisiert. Die raumzeitliche Lokalisierung sorgt zugleich für Identitätskriterien für Ereignisse: ‚Zwei' Ereignisse sind genau dann identisch, wenn sie am selben Ort und zur selben Zeit stattfinden.[17]

Ein Ereignis kann zahlreiche unterschiedliche Eigenschaften haben. Beispielsweise kann das Ereignis, auf das mit dem singulären Term ‚Thomas' Lauf am Rhein' referiert wird, die Eigenschaften haben, von Thomas ausgeführt worden zu sein, ein Lauf zu sein, 40 Minuten lang zu dauern und 10 km lang zu sein. Dieses Merkmal von Davidsons Ereignisbegriff ist relevant für die Formulierung

[15] Hierin folge ich Audi (2013: 2–3) und weiche von Heil (2003) ab.

[16] Davidson (1970: 138).

[17] Davidson (1969) zieht zunächst ein kausales Identitätskriterium diesem raumzeitlichen Identitätskriterium vor. In Reaktion auf Kritik durch Quine (1985) entschließt sich Davidson (1985: 248) jedoch zu diesem Identitätskriterium.

des Physikalismus, da hierdurch die Möglichkeit entsteht, dass ein und dasselbe Ereignis sowohl physische als auch nicht-physische Eigenschaften hat.

Jaegwon Kim schlägt einen anderen Ereignisbegriff vor: Er versteht unter Ereignissen „exemplifications by substances of properties at a time“[18]. Ein Ereignis besteht darin, dass ein bestimmtes Objekt zu einer bestimmten Zeit eine bestimmte Eigenschaft instantiiert. Thomas‘ Lauf z. B. könnte durch das Objekt *Thomas*, die Eigenschaft *Laufend* und die Zeit *9–9:40 Uhr* konstituiert sein. Kim würde das Ereignis dann in seiner Notation mit [Thomas, Laufend, 9–9:40 Uhr] bezeichnen. Allgemein werden Ereignisse in Kims Notation folgendermaßen bezeichnet: [o, E, t], wobei ‚o‘ für ein (für das Ereignis konstitutives) Objekt steht, ‚E‘ für eine konstitutive Eigenschaft und ‚t‘ für einen konstitutiven Zeitpunkt oder ein Zeitintervall.

Kim gibt folgende Existenzbedingung für Ereignisse an: Das Ereignis [o, E, t] existiert genau dann (oder findet genau dann statt), wenn das Objekt o zu Zeit t die Eigenschaft E instantiiert. Zudem gibt er eine Identitätsbedingung für Ereignisse an: Die Ereignisse [o, E, t] und [o‘, E‘, t ‘] sind genau dann identisch, wenn o und o‘ identisch sind, E und E‘ identisch sind und t und t‘ identisch sind.

Ebenso wie bei Davidson sind auch bei Kim Ereignisse nicht wiederholbar und raumzeitlich lokalisiert. Das liegt ganz einfach daran, dass sie unter anderem durch eine Zeit und ein Objekt bestimmt sind. Geht man davon aus, dass die konstitutive Zeit eines Ereignisses nicht wiederholbar ist, so ist auch das Ereignis nicht wiederholbar. Die räumliche Lokalisierung eines Ereignisses ist durch die räumliche Lokalisierung seines konstitutiven Objekts garantiert.

Anders als bei Davidson setzen die Identitätskriterien von Ereignissen bei Kim jedoch voraus, dass die konstitutiven Eigenschaften der ‚beiden‘ identischen Ereignisse ebenfalls identisch sind. Es ergeben sich daher eine ganze Reihe von unterschiedlichen Kim-Ereignissen, die am selben Ort und zur selben Zeit stattfinden. Thomas‘ Lauf, Thomas 12 km/h-Lauf, Thomas‘ Rhein-Lauf usw. können bei Kim als distinkte Ereignisse gezählt werden, da sie distinkte konstitutive Eigenschaften haben. Bei Davidson hingegen haben wir es hier mit nur einem – raumzeitlich individuierten – Ereignis zu tun. Dieser Unterschied wird oft durch eine Unterscheidung zwischen enger und weiter Individuation von Ereignissen eingefangen: Davidson verwendet eine weite Individuation von Ereignissen und zählt weniger Ereignisse, während Kim eine enge Individuation von Ereignissen verwendet und mehr Ereignisse zählt.[19]

[18] Kim (1976: 34).

[19] Dieser Unterschied ist unter Voraussetzung eines kausalen Identitätskriteriums für Davidson-Ereignisse weitaus weniger eindeutig. Die Frage nach der Identität vom Thomas‘

Ich werde mich im Folgenden Kims Auffassung von Ereignissen anschließen. Wenn ich also schlichtweg von Ereignissen rede, meine ich Ereignisse im Sinne Kims. Auch in diesem Fall aber gibt es Stellen, an denen der Wechsel zur Rede von Ereignissen im Sinne Davidsons systematische Konsequenzen hat. Besonders in den Abschnitten 2.6.1. und 2.6.2. gehe ich hierauf noch einmal ein.

Der Zusammenhang zwischen Eigenschaften und Ereignissen, der sich aus Kims Konzeption von Ereignissen ergibt, erlaubt auch einen einfachen Begriffswechsel in Teil 3 und Teil 4 dieser Arbeit. Da es in diesen Teilen darum geht, in welche Kausalverhältnisse mentale/spezialwissenschaftliche Entitäten eintreten und die Relata der Kausalrelation typischerweise als Ereignisse aufgefasst werden, ist dort meist die Rede von Ereignissen und nicht mehr von Eigenschaften. Die Kernthesen des nicht-reduktiven Physikalismus, die hier in Bezug auf Eigenschaften formuliert werden, übertragen sich jedoch unter Voraussetzung von Kims Konzeption sämtlich auf Ereignisse. Wenn spärliche mentale Eigenschaften existieren, dann existieren auch spärliche mentale Ereignisse (Realismus): Es handelt sich dann um solche Ereignisse, deren konstitutive Eigenschaft spärlich ist. Wenn mentale/spezialwissenschaftliche Eigenschaften multipel realisiert sind, kann man auch sagen, dass die zugehörigen mentalen Ereignisse multipel realisiert sind (Multiple Realisierung). Wenn mentale Eigenschaften nicht mit physischen Eigenschaften identisch sind, dann sind auch die zugehörigen mentalen Ereignisse nicht mit physischen Ereignissen identisch (Nicht-Identität). Und wenn mentale Eigenschaften von physischen Eigenschaften ontologisch abhängen, dann hängen auch die zugehörigen mentalen Ereignisse von physischen Ereignissen ontologisch ab (Ontologische Abhängigkeit).[20]

2.3 Physische Eigenschaften

Die Frage nach dem Verhältnis zwischen physischen und mentalen oder spezialwissenschaftlichen Eigenschaften gewinnt erst vor dem Hintergrund einer

Lauf und Thomas‘ Rhein-Lauf würde dann davon abhängen, ob Thomas‘ Lauf und Thomas‘ Rhein-Lauf dieselben Ursachen und Wirkungen haben. Gerade diese Frage ist aber entscheidend davon abhängig, welche Theorie der Kausalität vorausgesetzt wird.

[20] Teilweise müssten die hier behaupteten Zusammenhänge (zumindest in der Realismus-These und der These der ontologischen Abhängigkeit) genaugenommen um Bedingungen mit Blick auf die zugehörigen Objekte und Zeitpunkte ergänzt werden. Ich denke, dass die Zusammenhänge zumindest dann plausibel sind, wenn wir davon ausgehen, dass konstitutives Objekt und konstitutive Zeit der verglichenen Ereignisse identisch sind.

Charakterisierung physischer Eigenschaften Gehalt. Auch die These des nicht-reduktiven Physikalismus ist daher auf ein vorläufiges Verständnis physischer Eigenschaften angewiesen. Schließlich wird durch eine Charakterisierung physischer Eigenschaften erst festgelegt, von welcher Art Eigenschaften mentale und spezialwissenschaftliche Eigenschaften ontologisch abhängig sein sollen.[21]

Die Charakterisierung physischer Eigenschaften wirft jedoch einige Probleme auf, die in den folgenden Abschnitten diskutiert werden sollen. Hierfür führe ich zunächst zwei grundsätzliche Unterscheidungen ein: Eine Unterscheidung zwischen einem primären und einem sekundären Sinn physischer Eigenschaften (Abschnitt 2.3.1) und eine Unterscheidung zwischen einem engen und einem weiten Sinn physischer Eigenschaften (Abschnitt 2.3.2). Daraufhin gehe ich auf objektbasierte, theoriebasierte und negative Konzeptionen physischer Eigenschaften ein (Abschnitt 2.3.3) und diskutiere ein Problem für theoriebasierte Konzeptionen (Abschnitte 2.3.4 und 2.3.5).

2.3.1 Der primäre und der sekundäre Sinn physischer Eigenschaften

Die Auszeichnung einer Eigenschaft als *physisch im primären Sinne* bedeutet zunächst, dass die entsprechende Eigenschaft eine (noch zu spezifizierende) Bedingung erfüllt, die zum Beispiel in einer bestimmten Beziehung zu physischen Objekten, zu physikalischen Theorien oder in einer Abgrenzung zu mentalen Eigenschaften bestehen kann. Für den nicht-reduktiven Physikalismus wird auf diese Weise die Klasse der Eigenschaften herausgegriffen, die als ontologische Basis für alle weiteren Eigenschaften dienen sollen. Laut dem nicht-reduktiven Physikalismus sind physische Eigenschaften im primären Sinne also Eigenschaften, die eine bestimmte Bedingung erfüllen und die als ontologische Basis aller weiteren Eigenschaften dienen.

Neben diesem primären Sinn physischer Eigenschaften finden sich in der Literatur häufig Verwendungsweisen, die den Begriff des Physischen auch auf solche

[21] Die Relevanz eines angemessenen Verständnisses physischer Eigenschaften für den Gehalt des Physikalismus wird u. a. von Barbara Montero (1999: 183) betont: „As I see it, there is little use in arguing about whether the mind is physical, or whether mental properties are physical properties – questions many take to be central to the mind-body problem – unless we have at least some understanding of what it means to be physical."

Eigenschaften ausweiten, die in einer hinreichend starken Beziehung ontologischer Abhängigkeit zu primär physischen Eigenschaften stehen.[22] Eine solche Verwendungsweise finden wir etwa in der folgenden Passage bei Frank Jackson:

> It is undeniable that the physical, chemical and biological sciences have provided a great deal of information about the world we live in and about ourselves. I will use the label ‚physical information' for this kind of information, and also for information that automatically comes along with it.[23]

Jackson deutet hier zunächst eine Bedingung an physische Eigenschaften[24] an, die diese durch ihr Vorkommen in bestimmten wissenschaftlichen Theorien bestimmt. Dann weitet er den Begriff auf Eigenschaften aus, die automatisch mit Eigenschaften *einhergehen*, die diese Bedingung erfüllen. Manche Eigenschaften gewinnen ihren Status als physische Eigenschaften demzufolge dadurch, dass sie in einer bestimmten Beziehung zu primär physischen Eigenschaften stehen. Solche Eigenschaften sind zwar nicht (unbedingt) mit primär physischen Eigenschaften identisch, können aber dennoch in einem abgeleiteten Sinn als ‚physisch' bezeichnet werden. Es ergibt sich so ein weiterer Begriff von ‚physisch', der nicht nur auf primär physische Eigenschaften zutrifft, sondern auch auf solche Eigenschaften, die in einer geeigneten Beziehung zu primär physischen Eigenschaften stehen. Ich spreche in diesem Zusammenhang von sekundär physischen Eigenschaften.

Diese Unterscheidung zwischen primär und sekundär physischen Eigenschaften kann auf unterschiedliche Weise spezifiziert werden. Hier hängt alles davon ab, welche Beziehungen zu primär physischen Eigenschaften eine Eigenschaft als sekundär physisch auszeichnen. Typischerweise wird hier genau jene Beziehung ontologischer Abhängigkeit herangezogen, die nicht-reduktive Physikalist*innen für sich beanspruchen. Eine Eigenschaft ist demnach sekundär physisch, wenn sie entweder primär physisch ist oder in der Beziehung der starken ontologischen Abhängigkeit zu primär physischen Eigenschaften steht.[25]

Reduktive und nicht-reduktive Physikalist*innen sind sich dann darin einig, dass alle Eigenschaften – inklusive mentaler Eigenschaften – sekundär physisch sind. Denn die im nicht-reduktiven Physikalismus vertretene Kombination aus der

---

[22] Vgl. für die Unterscheidung zwischen primär physischen und sekundär physischen Eigenschaften auch Dowell (2006: 1).

[23] Jackson (1982: 127).

[24] Ich übertrage die Redeweise von Informationen im Zitat hier auf die Redeweise von Eigenschaften.

[25] Vgl. z. B. Zhong (2011: 135–136).

These der Nicht-Identität und der These der ontologischen Abhängigkeit lässt sich nur unter der Annahme plausibilisieren, dass in beiden Thesen der primäre Sinn physischer Eigenschaften ausschlaggebend ist: Tatsächlich folgt aus der These der ontologischen Abhängigkeit, dass mentale und spezialwissenschaftliche Eigenschaften mit sekundär physischen Eigenschaften identisch sind. Die These der Nicht-Identität muss daher in jedem Falle so verstanden werden, dass sie sich auf primär physische Eigenschaften bezieht.

Der sekundäre Sinn physischer Eigenschaften ist daher in Kontexten, in denen es um eine Abgrenzung zwischen reduktivem und nicht-reduktivem Physikalismus geht, eher wenig hilfreich. Reduktive und nicht-reduktive Physikalist*innen sind sich darin einig, dass mentale und spezialwissenschaftliche Eigenschaften sekundär physisch sind. Ihre Differenz ergibt sich aus der Frage, ob mentale und spezialwissenschaftliche Eigenschaften primär physisch sind. Stattdessen sollte die Verwendung des Begriffs sekundär physischer Eigenschaften in der Regel auf Kontexte beschränkt werden, in denen es um die Gegenüberstellung von physikalistischen und nicht-physikalistischen Positionen geht.

Vor diesem Hintergrund ist es sinnvoll, den Begriff des Physischen auf primär physische Eigenschaften zu beschränken. Wenn ich im Folgenden also ohne Qualifikation von ‚physischen Eigenschaften' spreche, meine ich damit immer primär physische Eigenschaften.

2.3.2 Physisch$_{eng}$ und physisch$_{weit}$

Der Begriff der sekundär physischen Eigenschaften ist weiter als der Begriff der primär physischen Eigenschaften. Manche Autor*innen setzen die Unterscheidung daher gleich mit einer Unterscheidung zwischen physischen Eigenschaften im engen Sinne und physischen Eigenschaften im weiten Sinne.[26] Eine solche Unterscheidung zwischen mehr oder weniger engen Bestimmungen physischer Eigenschaften lässt sich aber sinnvoll auch hinsichtlich des Begriffs primär physischer Eigenschaften ziehen.

Im obigen Zitat bestimmt Jackson primär physische Eigenschaften unter Bezug auf „die physikalischen, chemischen und biologischen Wissenschaften"[27] und verwendet damit einen relativ weiten Begriff primär physischer Eigenschaften. In Diskussionen des Problems der mentalen Verursachung werden weiterhin

---

[26] Vgl. z. B. Zhong (2011: 134–135), Rabin (2020: 7–8).

[27] Jackson (1982: 127).

neuronale Eigenschaften häufig als Paradigmen (primär) physischer Eigenschaften angeführt.[28] Der weite Sinn primär physischer Eigenschaften ist also dadurch ausgezeichnet, dass Eigenschaften, die von den Prädikaten der Physik oder einer ganzen Reihe unterschiedlicher Spezialwissenschaften bezeichnet werden, als primär physisch ausgezeichnet werden. Ein engerer Sinn primär physischer Eigenschaften würde hingegen nur die Eigenschaften, die von Prädikaten physikalischer Theorien bezeichnet werden, als physisch auffassen. Da dieser Unterschied für verschiedene Aspekte der Argumentation dieser Arbeit relevant ist, führe ich die Ausdrücke ‚$\text{physisch}_{\text{eng}}$' und ‚$\text{physisch}_{\text{weit}}$' ein, um der geschilderten Unterscheidung Rechnung zu tragen.

Sollte die Formulierung des nicht-reduktiven Physikalismus also auf einem engen oder einem weiten Sinn physischer Eigenschaften beruhen? Es gibt zumindest zwei Überlegungen, die dafür sprechen, physische Eigenschaften in den Kernthesen des nicht-reduktiven Physikalismus als $\text{physische}_{\text{eng}}$ Eigenschaften aufzufassen:

Allgemein wird der Physikalismus umso informativer und zugleich gewagter sowie sparsamer, je enger der Begriff primär physischer Eigenschaften gefasst wird.[29] Da der Physikalismus mit dem Anspruch auftritt, eine sparsame Ontologie zu sein, sollten primär physische Eigenschaften daher möglichst eng bestimmt werden.

Eine weite Bestimmung primär physischer Eigenschaften hat zudem den Nachteil, dass die daraus resultierende Version des Physikalismus die Beziehungen zwischen $\text{physischen}_{\text{weit}}$ Eigenschaften unterschiedlicher Art offenlässt: Unter Voraussetzung eines weiten Begriffs physischer Eigenschaften macht der Physikalismus beispielsweise keine Aussagen über das Verhältnis zwischen neuronalen und physikalischen Eigenschaften und lässt daher die Möglichkeit offen, dass neuronale Eigenschaften in einem starken Sinne emergent sind gegenüber den Eigenschaften, die Gegenstand physikalischer Theorien sind.

Im nicht-reduktiven Physikalismus steht nicht nur das Verhältnis zwischen mentalen und physischen Eigenschaften zur Debatte, sondern auch das Verhältnis zwischen spezialwissenschaftlichen und physischen Eigenschaften. Der nicht-reduktive Physikalismus enthält daher durchaus auch eine Ablehnung der These, dass neuronale Eigenschaften (als eine Klasse spezialwissenschaftlicher Eigenschaften) nicht emergent gegenüber den Eigenschaften sind, die in physikalischen

---

[28] Häufig bleibt jedoch unbestimmt, ob neurologische Eigenschaften als Paradigmen primär physischer Eigenschaften aufgefasst werden, oder ob sie als offenbare Fälle sekundär physischer Eigenschaften gesehen werden.

[29] Vgl. auch Poland (1994: 114).

Theorien behandelt werden. Versteht man die Kernthesen des nicht-reduktiven Physikalismus aber so, dass sie sich auf physische$_{\text{weit}}$ Eigenschaften beziehen, wird eine solche These durch den nicht-reduktiven Physikalismus nicht impliziert. Man sollte die Kernthesen des nicht-reduktiven Physikalismus daher so verstehen, dass sie sich auf physische$_{\text{eng}}$ Eigenschaften beziehen.

Eine andere Frage ist es, *wie eng* der Bereich physischer$_{\text{eng}}$ Eigenschaften aufgefasst werden sollte. Es finden sich in der Literatur gelegentlich Verwendungsweisen des Begriffs physischer Eigenschaften, die den Bereich physischer$_{\text{eng}}$ Eigenschaften auf *mikrophysikalische* Eigenschaften beschränken.[30] Damit würden bereits Eigenschaften wie Härte oder Temperatur als nicht-physisch$_{\text{eng}}$ klassifiziert werden. Entsprechend könnte der nicht-reduktive Physikalismus, der sich auf eine solche Bestimmung physischer$_{\text{eng}}$ Eigenschaften bezieht, daran scheitern, dass makroskopische Eigenschaften wie Härte oder Temperatur nicht in einem hinreichend starken Sinne von mikrophysikalischen Eigenschaften ontologisch abhängen. Der nicht-reduktive Physikalismus wäre nach diesem Verständnis nicht nur auf Annahmen über die Beziehung zwischen mentalen/spezialwissenschaftlichen Eigenschaften und den Eigenschaften, die Gegenstand physikalischer Theorien sind, festgelegt. Er wäre darüber hinaus auch auf Annahmen über die Beziehung zwischen makroskopischen physikalischen Eigenschaften und mikrophysikalischen Eigenschaften festgelegt.

Es scheint jedoch sinnvoll, den nicht-reduktiven Physikalismus von einer Festlegung auf einen *Mikrophysikalismus*, demzufolge alles letztlich *mikrophysikalisch* ist, freizuhalten.[31] Sollte sich demnach herausstellen, dass – wie Andreas Hüttemann argumentiert – makroskopische physikalische Phänomene nicht einseitig durch mikrophysikalische Phänomene determiniert sind, sondern vielmehr eine wechselseitige Abhängigkeit vorliegt, würde dies nicht gegen den Physikalismus sprechen. Stattdessen ist hierdurch nur der Mikrophysikalismus in Zweifel gezogen.

Es ergeben sich so die folgenden Einschränkungen an die Extension des für den nicht-reduktiven Physikalismus ausschlaggebenden Begriffs physischer$_{\text{eng}}$ Eigenschaften: Physische$_{\text{eng}}$ Eigenschaften sollten so bestimmt werden, dass mentale und spezialwissenschaftliche Eigenschaften nicht (per definitionem) als physische$_{\text{eng}}$ Eigenschaften zählen. Zugleich sollten sowohl mikrophysikalische Eigenschaften als auch makroskopische Eigenschaften, die von physikalischen

[30] Vgl. z. B. Chalmers (1996: 30–31) und Pettit (1993).

[31] Für eine Kritik am Mikrophysikalismus siehe Hüttemann (2004). Zum Verhältnis zwischen Mikrophysikalismus und Physikalismus vgl. Kapitel 8 in Hüttemann (2004), Hüttemann und Papineau (2005) und Papineau (2008).

Theorien behandelt werden, als physische$_{eng}$ Eigenschaften zählen. Diese extensionalen Einschränkungen lassen jedoch noch offen, wie die relevante Klasse von Eigenschaften herausgegriffen werden kann – welche Kriterien eine Eigenschaft also als physisch zählen lassen. Dies ist Gegenstand des nächsten Abschnitts.

2.3.3 Objektbasierte, theoriebasierte und negative Konzeptionen physischer Eigenschaften

Objektbasierte Konzeptionen physischer Eigenschaften gehen davon aus, dass wir über ein vorgängiges Paradigma von physischen Objekten verfügen: Physische Objekte sind zum Beispiel Tische, Waschmaschinen oder Steine – ‚mittelgroße Exemplare trockener Güter'[32]. Solche Objekte sind durch eine Reihe paradigmatischer Eigenschaften ausgezeichnet, wie ihre räumliche Ausdehnung, Größe und Form, Beweglichkeit und Festigkeit. Auf der Grundlage dieser Paradigmen von physischen Objekten und ihren Eigenschaften kommt Daniel Stoljar zur folgenden Definition physischer Objekte:

> x is a physical object if and only if x has (or has enough of) the following properties: it has size, shape, extension in space, the capacity to move and be moved, the capacity to undergo various processes such as bending, breaking, and burning, and perhaps most importantly it has solidity or bulk – that is, it is intrinsically such that it resists or would resist pressure from other physical objects, for example, pressure from human bodies.[33]

Ausgehend von einem solchen intuitiven Verständnis physischer Objekte könnte man versuchen, zu einer geeigneten Charakterisierung physischer Eigenschaften zu gelangen. Der direkteste Weg würde dabei darin bestehen, einfach all jene Eigenschaften als physisch zu zählen, die von physischen Objekten (oder Teilen physischer Objekte) instantiiert werden:[34]

> **(Ph-O)** Eine Eigenschaft P ist genau dann physisch, wenn gilt: P wird von einem physischen Objekt oder einem Teil eines physischen Objekts instantiiert.

[32] Diese verbreitete Redewendung geht auf Austin (1962: 8) zurück.

[33] Stoljar (2010: 52).

[34] Stoljar (2010: 53).

Im gegebenen dialektischen Kontext ist dies jedoch keine gute Idee: Auch menschliche Körper haben schließlich die genannten Eigenschaften und sind daher physische Objekte. Zugleich instantiieren sie aber mentale Eigenschaften. Mentale Eigenschaften würden daher *per definitionem* als physisch zählen, insofern sie von physischen Objekten instantiiert werden. Eine offene Diskussion über das Verhältnis zwischen mentalen und physischen Eigenschaften würde sich damit erübrigen. Insbesondere wäre die für den nicht-reduktiven Physikalismus ausschlaggebende These der Nicht-Identität analytisch falsch.

Negative Konzeptionen physischer Eigenschaften nehmen ihren Ausgang von bestimmten Kontrastklassen: Physische Eigenschaften werden dann rein negativ – das heißt in Abgrenzung zu ihrer Kontrastklasse – bestimmt. Negative Konzeptionen folgen also dem folgenden Schema:

> **(Ph-N)** Eine Eigenschaft P ist genau dann physisch, wenn gilt: P ist *nicht* X.

Die Kontrastklasse X kann dabei auf unterschiedliche Weisen spezifiziert werden. Im gegebenen Kontext ist insbesondere der Kontrast zu mentalen und spezialwissenschaftlichen Eigenschaften relevant. Es bietet sich also zunächst folgende Definition an:

> **(Ph-NM)** Eine Eigenschaft P ist genau dann physisch, wenn gilt: P ist *nicht* mental und *nicht* spezialwissenschaftlich.

Ein unmittelbares Problem mit dieser Definition besteht darin, dass sie Identitätstheorien begrifflich ausschließt: Mentale Eigenschaften können nicht mit physischen Eigenschaften identisch sein, wenn physische Eigenschaften *per definitionem* nicht mental sind.[35] Die These der Nicht-Identität müsste dann nicht – wie im nicht-reduktiven Physikalismus üblich – mit Verweis auf die These der multiplen Realisierung gerechtfertigt werden, sondern wäre analytisch wahr. Dies zeigt, dass (Ph-NM) keine geeignete Begriffsbestimmung physischer Eigenschaften ist.[36]

---

[35] Vgl. z. B. Stoljar (2010: 87).

[36] Eine negative Bestimmung physischer Eigenschaften wird gelegentlich mithilfe folgender Qualifikation verteidigt: Eine Eigenschaft *P* ist genau dann physisch, wenn gilt: *P* ist nicht *fundamental* mental. (Vgl. z. B. Wilson (2006)). Auch diese Bestimmung ist im Kontext des nicht-reduktiven Physikalismus jedoch ungeeignet: Nicht-reduktive Physikalist*innen stimmen zu, dass mentale Eigenschaften nicht *fundamental* mental und in diesem Sinne physisch sind. Die These der Nicht-Identität kann sich also nicht auf physische Eigenschaften in diesem Sinne beziehen.

Eine Modifikation der negativen Bestimmung physischer Eigenschaften[37], die diesem Problem aus dem Weg geht, ist die folgende:

> **(Ph-NM-M)** Eine Eigenschaft P ist genau dann physisch, wenn gilt: P kann durch nicht-mentales und nicht-spezialwissenschaftliches *Vokabular* bezeichnet werden.

Diese Definition erlaubt die Existenz von Eigenschaften, die sowohl physisch als auch mental sind: Denn ein und dieselbe Eigenschaft kann unter Umständen sowohl durch nicht-mentales als auch durch mentales Vokabular bezeichnet werden. Identitätstheorien werden durch (Ph-NM-M) also nicht begrifflich ausgeschlossen.

Ein anderes Problem mit negativen Konzeptionen physischer Eigenschaften besteht jedoch in einer geeigneten Bestimmung der Kontrastklasse. Erstens ist der Begriff einer physischen Eigenschaft durch eine negative Konzeption physischer Eigenschaften nur so klar bestimmt wie die herangezogene Kontrastklasse. Wenn physische Eigenschaften durch Abgrenzung zu mentalen und spezialwissenschaftlichen Eigenschaften bestimmt werden, sollte der Begriff einer mentalen/spezialwissenschaftlichen Eigenschaft also angemessen charakterisiert werden. Zweitens setzt die negative Konzeption voraus, dass die Menge aller Eigenschaften erschöpfend in die Kontrastklasse und die Klasse physischer Eigenschaften eingeteilt werden kann. Eventuell gibt es jedoch Eigenschaften, die intuitiv weder der Menge der physischen Eigenschaften noch der Menge der mentalen/spezialwissenschaftlichen Eigenschaften angehören. In Frage kämen hier etwa mathematische oder semantische Eigenschaften. Entsprechend müsste die Kontrastklasse für die Bestimmung physischer Eigenschaften dann erweitert werden. Es ist jedoch keinesfalls klar, dass man auf diese Weise je zu einer geeigneten Kontrastklasse gerät. Zudem stellt sich die Frage, ob wir auf diese Weise ein *aufschlussreiches Kriterium* für den Begriff physischer Eigenschaften gewinnen könnten. Eine Auflistung von Eigenschaften, die nicht unter den Begriff einer physischen Eigenschaft fallen, klärt noch nicht darüber auf, *warum* die verbleibenden Eigenschaften unter den Begriff physischer Eigenschaften fallen. Durch eine erschöpfende Auflistung von Dingen, die keine Bäume sind, habe ich noch keine wirklich aufschlussreiche Bestimmung des Begriffs ‚Baum' gegeben.[38]

---

[37] Vgl. Papineau (2002: 41–44) und Papineau (2008: 131). Ich danke Leonard Dung für einen entsprechenden Hinweis.

[38] Vgl. Stoljar (2010: 87–88).

Der vielversprechendste Ansatz zur Charakterisierung physischer Eigenschaften sind nach meiner Auffassung theoriebasierte Konzeptionen. Theoriebasierte Konzeptionen bestimmen physische Eigenschaften als solche Eigenschaften, die durch Prädikate bestimmter Theorien bezeichnet werden. Diese Idee kann in folgendem Schema festgehalten werden:

(Ph-T) Eine Eigenschaft P ist genau dann physisch, wenn gilt: P wird durch ein Prädikat einer Theorie T bezeichnet.

Der Gehalt einer solchen Definition hängt wesentlich davon ab, wie die relevanten Theorien T bestimmt werden. Von welcher Theorie (oder welchen Theorien) sollten wir sagen, dass ihre Prädikate physische Eigenschaften bezeichnen, um einen geeigneten Begriff des Physischen zu erhalten?

Nach dem in Abschnitt 2.3.2 Gesagten ist es naheliegend, sich bei der Bestimmung des Begriffs physischer$_{eng}$ Eigenschaften allgemein auf physikalische Theorien zu beziehen, ohne aber bestimmte Theorien innerhalb der Physik gesondert herauszugreifen. Es ergibt sich so die folgende Definition physischer$_{eng}$ Eigenschaften:

(Ph$_{eng}$-T) Eine Eigenschaft P ist genau dann physisch$_{eng}$, wenn gilt: P wird durch ein Prädikat einer physikalischen Theorie bezeichnet.

Als physikalische Theorie gelten dabei sowohl das Standardmodell der Teilchenphysik und die allgemeine Relativitätstheorie oder ein Ansatz zur Vereinheitlichung dieser Theorien, als auch Theorien der Thermodynamik, der Festkörperphysik oder der Kosmologie. Beziehen sich die Kernthesen des nicht-reduktiven Physikalismus also auf so verstandene physische$_{eng}$ Eigenschaften, vermeiden sie eine Festlegung auf einen Mikrophysikalismus. Da spezialwissenschaftliche Theorien nicht in diese Definition physischer$_{eng}$ Eigenschaften mit einbezogen sind, ergibt sich zugleich eine offene und interessante Frage betreffend die Beziehung zwischen spezialwissenschaftlichen und physischen$_{eng}$ Eigenschaften.

Auch der Begriff physischer$_{weit}$ Eigenschaften lässt sich aber theorienbasiert definieren:

(Ph_{weit}-T) Eine Eigenschaft P ist genau dann $physisch_{weit}$, wenn gilt: P wird durch ein Prädikat einer physikalischen Theorie oder einer spezialwissenschaftlichen (z. B. chemischen, biologischen, neurowissenschaftlichen, usw.) Theorie bezeichnet.[39]

Wie in Abschnitt 2.3.2. schon gesagt, sollte der nicht-reduktive Physikalismus nicht so verstanden werden, dass er sich auf so verstandene $physische_{weit}$ Eigenschaften bezieht. Denn dies würde dazu führen, dass die These der Nicht-Identität (‚Mentale und spezialwissenschaftliche Eigenschaften sind nicht mit physischen Eigenschaften identisch') trivialerweise falsch wird. Stattdessen ist der durch die Definition (Ph_{eng}-T) gegebene Begriff physischer Eigenschaften ausschlaggebend für den nicht-reduktiven Physikalismus.

Dennoch werde ich auf die Unterscheidung zwischen $physischen_{eng}$ und $physischen_{weit}$ Eigenschaften an der einen oder anderen Stelle dieser Arbeit noch einmal eingehen. Gemeint sind dabei durchgehend die Begriffe, die durch die Definition (Ph_{eng}-T) bzw. (Ph_{weit}-T) gegeben sind.

2.3.4 Hempels Dilemma

Carl Gustav Hempel schildert in der folgenden Passage ein Problem für theoriebasierte Formulierungen des Physikalismus, das unter dem Namen ‚Hempels Dilemma' Einzug in die Debatte gefunden hat:

> [T]he physicalistic claim that the language of physics can serve as a unitary language of science is inherently obscure: The language of *what* physics is meant? Surely not that of, say, 18th century physics; for it contains terms like 'caloric fluid', whose use is governed by theoretical assumptions now thought false. Nor can the language of contemporary physics claim the role of unitary language, since it will no doubt undergo further changes, too. The thesis of physicalism would seem to require a language in which a true theory of all physical phenomena can be formulated. But it is quite

[39] Es stellt sich hier jedoch die Frage, aufgrund welcher Kriterien eine Spezialwissenschaft in die Liste der in (Ph_{weit}-T) erwähnten Wissenschaften aufgenommen wird und ob aufgrund dieser Kriterien eine grundsätzliche Abgrenzung dieser Wissenschaften zur Psychologie möglich ist. Wenn eine solche prinzipiengeleitete Abgrenzung nicht möglich ist, wird die Psychologie also entweder willkürlich nicht in die Liste aufgenommen, oder aber sie *wird* in die Liste aufgenommen und mentale Eigenschaften zählen als physisch. Die letztgenannte Option würde jedoch wiederum dazu führen, dass der Physikalismus trivial wird. Vgl. hierzu auch Crane und Mellor (1990).

> unclear what is to be understood here by a physical phenomenon, especially in the context of a doctrine that has taken a determinedly linguistic turn.[40]

Hempel formuliert das Problem in dieser Passage nicht mit Blick auf den nicht-reduktiven Physikalismus, sondern mit Blick auf die Idee, dass es eine physikalische Universalsprache gibt, in die sämtliche empirisch gehaltvollen Sätze übersetzt werden können. Entsprechend bezieht sich seine Formulierung des Problems auch nicht auf die Charakterisierung physischer Eigenschaften, sondern auf die Charakterisierung der physikalischen Universalsprache.

Offenbar überträgt sich das Problem aber auf alle theoriebasierten Konzeptionen physischer Eigenschaften: Sollen physische Eigenschaften solche Eigenschaften sein, die durch Prädikate *gegenwärtiger* physikalischer Theorien bezeichnet werden, oder sollen sie solche Eigenschaften sein, die durch Prädikate einer *zukünftigen, idealen* physikalischen Theorie bezeichnet werden? Gegenwärtige Theorien kommen laut Hempel nicht in Frage, weil sie wahrscheinlich falsch und unvollständig sind. Zukünftige, ideale Theorien haben hingegen eine bessere Chance auf Wahrheit, sind jedoch bisher unbekannt. In beiden Fällen erhalten wir keine zufriedenstellende Charakterisierung physischer Eigenschaften.[41] Zur Verdeutlichung gehe ich auf beide Hörner des Dilemmas noch einmal etwas genauer ein:

Das erste Horn des Dilemmas besteht darin, die gegenwärtige Physik als Maßstab für die Bestimmung physischer Eigenschaften zu verwenden. Das Problem, das sich aus dieser Option ergibt, hebt mit der These an, dass die gegenwärtige Physik wahrscheinlich falsch und unvollständig ist. Diese These kann einerseits unter Verweis auf die spezifische Beschaffenheit der gegenwärtigen Physik – beispielsweise ihre Uneinheitlichkeit – gerechtfertigt werden. Andererseits ließe sich das aus der Debatte um den wissenschaftlichen Realismus bekannte Argument der pessimistischen Metainduktion[42] ins Spiel bringen: Da vergangene physikalische Theorien sich regelmäßig als falsch herausgestellt haben und die gegenwärtige Physik diesen vergangenen Theorien in relevanter Hinsicht ähnelt, wird sich wahrscheinlich auch die gegenwärtige Physik als falsch herausstellen.

Warum kann eine falsche oder unvollständige Physik nicht zu einer angemessenen Bestimmung physischer Eigenschaften führen? Erstens ist die Idee, dass es unentdeckte physische Eigenschaften gibt, nicht inkonsistent oder offenbar absurd. Wenn physische Eigenschaften aber über die gegenwärtige Physik

[40] Hempel (1980: 194–195).

[41] Vgl. auch Dowell (2006).

[42] Vgl. Laudan (1981).

definiert werden, sind solche unentdeckten physischen Eigenschaften begrifflich ausgeschlossen. Zweitens hat die physikalistische Grundthese, dass alle Eigenschaften letztlich physisch sind, unter der Voraussetzung, dass nur die Eigenschaften einer falschen oder unvollständigen Physik als physisch zählen, keine hohe Ausgangsplausibilität. Wenn es unentdeckte physische Eigenschaften gibt, ist davon auszugehen, dass sie auch eine unverzichtbare Rolle in der ontologischen Basis für alle weiteren Eigenschaften spielen. Zudem wären die unentdeckten physischen Eigenschaften nach der nun vorausgesetzten Konzeption physischer Eigenschaften selbst nicht physisch. Auch wären sie vermutlich nicht sekundär physisch im Sinne einer hinreichend starken ontologischen Abhängigkeit von primär physischen Eigenschaften. Der Physikalismus wäre unter Voraussetzung dieser Konzeption physischer Eigenschaften also selbst wahrscheinlich falsch. Zudem wäre er aus den falschen Gründen falsch: Wenn eine konservative Erweiterung der gegenwärtigen Physik die ontologische Basis für alle weiteren Eigenschaften vollständig charakterisieren würde, wäre der Physikalismus nach allgemeinem Verständnis wahr. Die aus der gegenwärtigen Theorien-Konzeption resultierende Version des Physikalismus wäre aber falsch. Die gegenwärtige Theorien-Konzeption würde also außerdem zu einer Formulierung des Physikalismus führen, die den etablierten Sprachgebrauch des Ausdrucks ‚Physikalismus' nicht korrekt einfängt.[43]

Das zweite Horn des Dilemmas besteht darin, eine zukünftige, ideale Physik als Maßstab für die Bestimmung physischer Eigenschaften heranzuziehen. Aus dieser Option ergeben sich zwei Probleme: Erstens ist der Gehalt einer idealen Physik unbestimmt und bestenfalls Gegenstand wilder Spekulation. Diese Unbestimmtheit überträgt sich auf den Begriff physischer Eigenschaften und auf die resultierende Formulierung des Physikalismus. Der Physikalismus wird so zu einer *völlig unklaren* Position. Zweitens wird die These vertreten, dass eine ideale Physik als eine ‚theory of everything' ihrer Natur nach sämtliche Eigenschaften, die zu einer vollkommenen Erklärung aller Phänomene notwendig sind, integrieren würde.[44] Wenn die physikalistische Grundthese, dass alle Eigenschaften letztlich physisch sind, also bei konservativen Erweiterungen der gegenwärtigen Physik nicht plausibel wird, werden die widerspenstigen, zum Beispiel mentalen, Eigenschaften dieser Auffassung zufolge schlicht in die ideale Physik integriert. Sie würden dann selbst als physisch zählen. Diese Überlegung legt nahe, dass

[43] Vgl. Stoljar (2010).

[44] Vgl. Chomsky (1972: 98) und (1980:. 5–6).

der Physikalismus unter einer Konzeption physischer Eigenschaften, die sich auf eine ideale Physik beruft, zu einer *trivialen* Position wird.[45]

Zusammenfassend kann also die gegenwärtige Physik nicht als Grundlage einer Konzeption physischer Eigenschaften dienen, weil die daraus resultierende Formulierung des Physikalismus wahrscheinlich falsch wäre und zudem nicht mit dem etablierten Sprachgebrauch des Physikalismus-Begriffs übereinstimmen würde. Die ideale Physik kann hingegen nicht als Grundlage für eine Konzeption physischer Eigenschaften herhalten, weil der Physikalismus so zu einer völlig unklaren oder trivialen These werden würde.

2.3.5 Wege aus Hempels Dilemma

Hempels Dilemma scheint zunächst ein ernstzunehmendes Problem für theoriebasierte Konzeptionen physischer Eigenschaften darzustellen. Dennoch lassen sich einige Überlegungen anstellen, die das Problem zumindest abschwächen:

Zunächst einmal beruht Hempels Dilemma auf einer falschen Dichotomie: Physische Eigenschaften müssen nicht entweder unter Bezug auf gegenwärtige oder in Bezug auf ideale Theorien bestimmt werden. Stattdessen ist es auch möglich, von einer allgemeinen Charakterisierung physikalischer Theorien auszugehen, die sowohl die gegenwärtige Physik als auch bestimmte ihrer Vorgänger und Nachfolger sowie zukünftige ideale Theorien als physikalische Theorien zählt. Dies ist ohnehin wünschenswert, da der Physikalismus begrifflich nicht auf eine detaillierte Konzeption physischer Eigenschaften, die sich an einer bestimmten physikalischen Theorie orientiert, festgelegt zu sein scheint: So sollte eine Welt, in der sämtliche Phänomene erschöpfend durch die Newton'sche Physik erklärt werden können, ebenso als physikalistisch zählen, wie eine Welt, in der die gegenwärtige Physik eine erschöpfende Erklärung sämtlicher Phänomene ermöglicht. Die Newton'sche Welt würde aber weder bei einer gegenwärtigen Theorienkonzeption noch bei einer idealen Theorienkonzeption als physikalistisch zählen. Dementsprechend sind beide Konzeptionen letztlich nicht angemessen.[46]

Das Problem besteht also weniger in der Entscheidung zwischen gegenwärtiger oder idealer Physik, als vielmehr in der Frage, was genau *im Allgemeinen* eine physikalische Theorie ausmacht. Der Begriff einer physischen Eigenschaft sollte an *keine spezifische, detailliert ausgearbeitete* physikalische Theorie definitorisch gebunden werden. Die Wahrheit des Physikalismus steht und fällt nicht

---

[45] Vgl. Dowell (2006).

[46] Vgl. Stoljar (2010: 101).

mit der Wahrheit einer bestimmten physikalischen Theorie. Der ausschlaggebende Sinn physischer Eigenschaften sollte entsprechend neutral gegenüber spezifischen physikalischen Theorien sein.

Das zweite Horn von Hempels Dilemma verweist jedoch bereits darauf, dass die Bestimmung des Begriffs einer physikalischen Theorie eine Herausforderung ist: Der Begriff darf nicht unbestimmt gelassen werden und er darf nicht so liberal bestimmt werden, dass eine Theorie, die unerklärte mentale Eigenschaften postuliert, als physikalisch zählt.

Allerdings scheint die Situation nicht so aussichtslos zu sein, wie es in Hempels Dilemma suggeriert wird: So kann eine Bestimmung des Begriffs einer physikalischen Theorie zumindest von den folgenden drei Ressourcen zehren: Erstens kennen wir paradigmatische Beispiele physikalischer Theorien, wie Newtons Mechanik, die allgemeine Relativitätstheorie und die Quantenmechanik. Physikalische Theorien sollten diesen paradigmatischen Beispielen ähneln.[47] Zweitens müssen physikalische Theorien allgemeine Standards der Wissenschaftlichkeit erfüllen: Sie sind empirisch prüfbar und zu einem gewissen Grade bestätigt, sie erklären empirische Phänomene auf der Grundlage von Gesetzen und sie postulieren theoretische Entitäten. Diese Eigenschaften teilt die gegenwärtige Physik sowohl mit Newtons Mechanik als auch mit ihren hypothetischen Nachfolgern.[48] Drittens können Überlegungen zum spezifischen Gegenstandsbereich der Physik hinzutreten: So bestimmt etwa Daniel Stoljar physikalische Theorien als „a theory that a scientist advances in the course of trying to explain or describe ordinary physical objects, their distinctive properties, their constitution and behavior, and so on“.[49] Hier wäre die Idee also, physikalische Theorien über ihre Rückbindung an die Erklärung des Verhaltens makroskopischer Objekte (wie Steine, Tische, Waschmaschinen oder Planeten) zu charakterisieren. Auf Grundlage dieser Ansatzpunkte ist zumindest eine Einschränkung des Begriffs einer physikalischen Theorie und somit eine Abschwächung von Hempels Dilemma möglich.

Trotz der verbleibenden Unschärfe hinsichtlich des Begriffs physischer Eigenschaften schließe ich mich im Folgenden der theoriebasierten Konzeption physischer Eigenschaften an. Der ausschlaggebende Sinn physischer Eigenschaften ist also durch die unter 2.3.3. formulierten Definitionen $\text{physischer}_{\text{eng}}$ Eigenschaften gegeben.

---

[47] Vgl. z. B. Ravenscroft (1997: 423–428).

[48] Vgl. Dowell (2006).

[49] Stoljar (2010: 73).

2.4 Die nicht-reduktiven Kernthesen des nicht-reduktiven Physikalismus

In Abschnitt 2.1.4. habe ich die nicht-reduktiven Kernthesen des nicht-reduktiven Physikalismus wie folgt bestimmt:

> **Realismus**: Der Realismus in Bezug auf mentale und spezialwissenschaftliche Aussagen und Eigenschaften ist wahr.
> **Nicht-Identität**: Mentale und spezialwissenschaftliche Eigenschaften sind nicht mit physischen Eigenschaften identisch.
> **Multiple Realisierung**: Mentale und spezialwissenschaftliche Eigenschaften sind auf physischer Ebene multipel realisiert.

Diese drei Thesen sind wechselseitig voneinander abhängig und sollten als Paket verstanden werden. Genauer gesagt ist der Realismus in Bezug auf mentale und spezialwissenschaftliche Eigenschaften Voraussetzung für die Thesen der Nicht-Identität und der multiplen Realisierung. Zudem stützt die These der multiplen Realisierung die These der Nicht-Identität.

Ziel dieses Abschnitts ist es, die genannten Kernthesen und ihre Beziehung zueinander genauer zu erläutern. Im Folgenden gehe ich unter 2.4.1. zunächst auf die genauere Interpretation der These des Realismus ein, bevor ich mich unter 2.4.2. der These der Nicht-Identität und unter 2.4.3. der These der multiplen Realisierung widme.

2.4.1 Realismus

Im Folgenden unterscheide ich eine Reihe von zusammenhängenden Thesen, die häufig mit realistischen Positionen in Bezug auf mentale und spezialwissenschaftliche Eigenschaften in Verbindung gebracht werden. Diese Thesen können in drei Familien zergliedert werden: Semantische Thesen, epistemische Thesen und ontologische Thesen.[50] Ich gehe auf die drei Gruppen in der Reihenfolge ein.

Zu semantischen Thesen: Die semantischen Thesen des Realismus betreffen die Interpretation von mentalem und spezialwissenschaftlichem Diskurs. Wir können sie hier als Thesen über semantische Eigenschaften von mentalen und spezialwissenschaftlichen Aussagen verstehen. Die grundlegendste semantische These ist erst

[50] Vgl. auch Bartels (2007: 200–202).

einmal, dass solche Aussagen überhaupt wahrheitswertfähig sind: Aussagen wie ‚Susi glaubt, dass Paris in Frankreich ist' sind wahr oder falsch. Zweitens wird dabei ein im weitesten Sinne korrespondenztheoretischer Wahrheitsbegriff vorausgesetzt: Die Wahrheit oder Falschheit einer mentalen Aussage misst sich zumindest teilweise an der Beschaffenheit der Welt, nicht aber allein an der Bedeutung der Bestandteile der Aussage, an Konsens in einem idealen Diskurs, an der Kohärenz eines einbettenden Begriffssystems, oder an ihrer Nützlichkeit. Diese beiden Thesen verstehe ich als die minimalen Thesen eines semantischen Realismus: Mentale und spezialwissenschaftliche Aussagen sind wahrheitswertfähig und ihre Wahrheitswerte ergeben sich aus der Beschaffenheit der Welt.

Die Verpflichtung auf einen im weitesten Sinne korrespondenztheoretischen Wahrheitsbegriff lässt das genauere Verhältnis zwischen wahren Aussagen und den wahrmachenden Elementen der Welt weitgehend offen. Wenn man hier präziser werden will, eröffnen sich zumindest die folgenden beiden theoretischen Optionen:

Zum einen könnte man vertreten, dass die Wahrheit mentaler Aussagen kaum spezifischere Aussagen über die Beschaffenheit der Welt zulässt. Man könnte dann zum Beispiel aus der Aussage ‚Susi glaubt, dass Paris in Frankreich ist' nicht folgern, dass es eine (spärliche) mentale Eigenschaft von Susi gibt, die darin besteht, zu glauben, dass Paris in Frankreich ist, und die Susi mit anderen Personen mit derselben Überzeugung teilt. Aus der Aussage ‚Susi und Peter glauben beide, dass Paris in Frankreich ist' würde nicht folgen, dass es eine genuine Gemeinsamkeit zwischen Susi und Peter gibt. Stattdessen könnte die Aussage aufgrund von ganz unterschiedlichen (physischen$_{eng}$) Eigenschaften von Susi und Peter wahr sein. Zwar werden die Aussagen dann durch bestimmte Elemente der Welt wahrgemacht, die Beziehung zwischen den wahrmachenden Elementen der Welt und dem (intuitiven) Gehalt der Aussagen ist aber eher lose und undurchsichtig.[51]

---

[51] Heil (2003) könnte eine solche Position zugeschrieben werden. Allerdings ist er hier nicht ganz eindeutig: Er lehnt zwar explizit eine ‚Picture Theory of Language' ab, der zufolge aus Aussagen der im Text genannten Art unmittelbar auf entsprechende spärliche Eigenschaften geschlossen werden kann. Jedoch verwendet er den Eigenschaftsbegriff auf eine Weise, die exakte Ähnlichkeit zwischen den Objekten, die eine Eigenschaft teilen, verlangt, während ich hier auch dann vom Teilen einer Eigenschaft sprechen möchte, wenn die Objekte, die die Eigenschaft teilen, einander in der relevanten Hinsicht nicht-exakt aber genuin ähneln. Es bleibt unklar, ob Heil aus der im Text genannten Aussage nicht-exakte Ähnlichkeiten zwischen Susi und Peter folgern würde.

Ney (2016a) diskutiert zudem die Aussichten eines ‚nicht-skeptischen Anti-Realismus', der sich auf eine bestimmte Version der Grounding-Relation beruft. Diese Position enthält die hier angesprochene Idee einer korrespondenztheoretischen Wahrheit mentaler und spezialwissenschaftlicher Aussagen ohne ontologische Verpflichtungen auf spärliche mentale und spezialwissenschaftliche Eigenschaften. Vgl. Wilson (2016: 500–502) für Kritik an dieser

Zum anderen könnte man eine engere Verbindung zwischen mentalen Aussagen und den sie wahrmachenden Elementen der Welt verlangen. Aus der Aussage ‚Susi und Peter glauben beide, dass Paris in Frankreich ist' folgt dann nicht bloß, dass es irgendein wahrmachendes Element in der Welt gibt, sondern es folgt auch, dass es eine spärliche mentale Eigenschaft gibt, die Susi und Peter teilen. Der mentale Diskurs ist dann gewissermaßen näher an der Beschaffenheit der Welt, zugleich aber weniger ontologisch neutral.

Die Kombination der minimalen Thesen des semantischen Realismus mit der ersten soeben skizzierten Konzeption kann als *schwacher semantischer Realismus* bezeichnet werden: Mentale Aussagen sind in einem korrespondenztheoretischen Sinne wahr, wobei sie aber keine spezifischeren Schlüsse auf die Beschaffenheit ihrer Wahrmacher zulassen. Die Kombination der minimalen Thesen mit der zweiten Konzeption kann als *starker semantischer Realismus* gelten: Mentale Aussagen sind in einem korrespondenztheoretischen Sinne wahr, wobei ihre Wahrheit oder Falschheit auch von der Existenz mentaler Entitäten wie insbesondere spärlicher mentaler Eigenschaften abhängt.

Zu epistemischen Thesen: Die epistemischen Thesen des Realismus betreffen die Erkennbarkeit von mentalen und spezialwissenschaftlichen Wahrheiten. Hier lautet die zentrale Frage: Können wir erkennen, ob mentale und spezialwissenschaftliche Aussagen wahr oder falsch sind? Auch hier lässt sich zunächst einmal eine anspruchslose epistemische These aufstellen: Mentale und spezialwissenschaftliche Aussagen sind prinzipiell als wahr oder falsch erkennbar. Diese minimale epistemische These ist nicht sehr kontrovers und wird allenfalls von radikalen Skeptizist*innen bestritten. Es ist zu beachten, dass diese These auch einen *minimalen* semantischen Realismus voraussetzt, demzufolge mentale und spezialwissenschaftliche Aussagen wahrheitswertfähig sind. Da diese Version des epistemischen Realismus jedoch nicht zwangsläufig mit einer Festlegung auf einen im weiten Sinn korrespondenztheoretischen Wahrheitsbegriff einhergeht, wird hierdurch weder ein schwacher noch ein starker semantischer Realismus impliziert.[52]

Verwendung von Grounding sowie Morris (2019: 172–182) und Ney (2020: 467–469) für weitere Diskussion.

Auch der ‚Truthmaking Physicalism', wie er von Morris (2018) kritisch diskutiert wird, fällt in diese Klasse von Theorien. Hier soll eine Relation des Wahrmachens die korrespondenztheoretische Verbindung zwischen der Wahrheit von mentalen und spezialwissenschaftlichen Sätzen und den zugehörigen physischen Realitäten herstellen.

[52] Vgl. auch Bartels (2007: 202) für eine entsprechende Anmerkung zum epistemischen wissenschaftlichen Realismus.

Zweitens kann aber die deutlich stärkere These aufgestellt werden, dass viele mentale und spezialwissenschaftliche Aussagen, die wir tatsächlich aufstellen, als *Wissen* gelten können. Da Wissen Wahrheit impliziert, würde aus dieser These zugleich folgen, dass viele mentale und spezialwissenschaftliche Aussagen *wahr* sind. Auch diese These ist (nicht nur von nicht-reduktiven Physikalist*innen) weitgehend akzeptiert. Tatsächlich kann die Annahme der Wahrheit vieler mentaler und spezialwissenschaftlicher Aussagen als gemeinsamer Ausgangspunkt der Debatte zwischen reduktiven und nicht-reduktiven Physikalist*innen aufgefasst werden, der sie insbesondere von eliminativen Physikalist*innen unterscheidet.[53]

Zu ontologischen Thesen: Die ontologischen Thesen des Realismus betreffen die Existenz von mentalen und spezialwissenschaftlichen Entitäten. Die minimale ontologische These ist dabei die folgende, simple *Existenzthese*: Es gibt mentale und spezialwissenschaftliche Eigenschaften. Mit dem Realismus wird zweitens traditionell eine *Unabhängigkeitsthese* verbunden: Die Existenz mentaler und spezialwissenschaftlicher Eigenschaften ist geist- und sprachunabhängig.

Auch die Existenzthese kann eine recht anspruchslose und wenig kontroverse Deutung erfahren: Wenn Eigenschaften im Sinne von üppigen Eigenschaften (siehe Abschnitt 2.2.1) verstanden werden, lassen sich recht leicht auch solche Eigenschaften finden, die alle Personen, die glauben, dass Paris in Frankreich ist, ‚teilen': Wenn etwa jeder beliebigen Menge von Objekten eine Eigenschaft entspricht, folgt recht unmittelbar, dass es eine solche Eigenschaft gibt – denn dafür muss man nur eine entsprechende Menge finden. Zugleich müssen sich jedoch Objekte, die eine üppige Eigenschaft teilen, nicht wirklich in irgendeiner Hinsicht ähneln. Die These, dass es mentale und spezialwissenschaftliche Eigenschaften gibt, wird daher in aller Regel anspruchsvoller gedeutet: Es gibt mentale und spezialwissenschaftliche *spärliche* Eigenschaften.[54] Mit dieser These ist nun verbunden, dass die Welt echte Ähnlichkeiten zwischen Objekten in mentalen und spezialwissenschaftlichen Hinsichten enthält.

Mit der Existenzthese ist jedoch nicht verbunden, dass mentale und spezialwissenschaftliche Eigenschaften auch *fundamental* sind. Eine Möglichkeit, den Begriff der Fundamentalität zu explizieren, verweist auf ontologische Abhängigkeitsbeziehungen der Art, die in Abschnitt 2.5. dieses Kapitels behandelt wird. Fundamentale Eigenschaften können dann als Eigenschaften aufgefasst werden, die von keinen

[53] Man beachte aber, dass eliminative Physikalist*innen nicht die minimale These bestreiten, dass mentale Aussagen als wahr oder falsch erkennbar sind. Vielmehr setzt ihre Argumentation gerade voraus, dass wir sie als falsch erkennen können.

[54] Vgl. z. B. Loewer (2007: 245).

anderen Eigenschaften ontologisch abhängig sind.[55] Für nicht-reduktive Physikalist*innen ist es zentral, dass mentale und spezialwissenschaftliche Eigenschaften gerade nicht in diesem Sinne fundamental sind und dass sie dennoch mit echten Ähnlichkeiten in der Natur einhergehen. Es gibt laut nicht-reduktiven Physikalist*innen also nicht-fundamentale spärliche Eigenschaften.[56]

Die Unabhängigkeitsthese lässt in ihrer Anwendung auf mentale Eigenschaften eine Schwierigkeit entstehen: Mentale Eigenschaften sind in einem trivialen Sinne geistabhängig, da sie selbst den ‚Geist' konstituieren. Die Idee, die hier jedoch eingefangen werden soll und die auch in Bezug auf mentale Eigenschaften ihre Anwendung findet, ist die folgende: Mentale Eigenschaften verdanken ihre Existenz nicht dem Umstand, dass sie selbst mental repräsentiert werden und sie verdanken ihre Beschaffenheit nicht der Art und Weise, wie sie repräsentiert werden.[57] Zwar können mentale Eigenschaften selbst noch einmal Gegenstand mentaler Repräsentation sein. Sie sind dies jedoch nicht wesentlich und ihre Beschaffenheit ist (zumindest teilweise) unabhängig davon, ob sie repräsentiert werden: Susi kann auch glauben, dass Paris in Frankreich ist, wenn sie es niemandem erzählt und die Überzeugung selbst nicht reflektiert. Zumindest in diesem Sinne soll gelten: „minds are what they are independently of how we take them to be".[58]

Der nicht-reduktive Physikalismus ist zu einem guten Stück durch einen robusten mentalen und spezialwissenschaftlichen Realismus motiviert. Hierzu gehört nach meiner Auffassung zusammenfassend eine Festlegung auf die jeweils stärksten Versionen der soeben skizzierten realistischen Thesen:

Starker Semantischer Realismus: Mentale und spezialwissenschaftliche Aussagen sind in einem korrespondenztheoretischen Sinne wahr, wobei ihre Wahrheit oder Falschheit auch von der Existenz mentaler und spezialwissenschaftlicher Entitäten wie insbesondere spärlicher mentaler und spezialwissenschaftlicher Eigenschaften abhängt.
Epistemischer Realismus: Wir haben Wissen über viele mentale und spezialwissenschaftliche Aussagen. Viele mentale und spezialwissenschaftliche Aussagen sind wahr.

---

[55] Vgl. Kapitel 5 in Bennett (2017) und Abschnitt 2.5.8. dieser Arbeit.

[56] Vgl. auch die Diskussion von Schaffer (2004a) sowie Loewer (2007: 245).

[57] Vgl. Vinueza (2001) für eine Analyse von Geistunabhängigkeit, die in diese Richtung geht. Die Idee könnte auf ihre Grenzen stoßen, wenn sie mit einer ‚higher order'-Theorie des Bewusstseins (vgl. Carruthers (2016)) verbunden wird.

[58] Heil (2003: 11).

Ontologischer Realismus: Es gibt spärliche mentale und spezialwissenschaftliche Eigenschaften, deren Existenz sprach- und geistunabhängig ist.

Aus der Kombination aus semantischem und epistemischem Realismus ergibt sich dabei ein zumindest partielles Argument für den ontologischen Realismus: Denn wenn mentale und spezialwissenschaftliche Aussagen, wie der epistemische Realismus behauptet, zumindest teilweise wahr sind und ihre Wahrheit, wie der semantische Realismus behauptet, Implikationen für die Existenz mentaler und spezialwissenschaftlicher Eigenschaften hat, folgt daraus die Existenz mentaler und spezialwissenschaftlicher Eigenschaften.

2.4.2 Nicht-Identität

Die These der Nicht-Identität ist eine ontologische These über den Zusammenhang zwischen mentalen bzw. spezialwissenschaftlichen Eigenschaften und physischen Eigenschaften. Der ontologische Realismus über mentale und spezialwissenschaftliche Eigenschaften wird dabei vorausgesetzt. Entsprechend bezieht sich die These der Nicht-Identität auf *spärliche* mentale und spezialwissenschaftliche Eigenschaften.

Hinsichtlich des Bezugs der These der Nicht-Identität stellt sich eine Frage, die die in Abschnitt 2.3.2. eingeführte Unterscheidung zwischen $\text{physischen}_{\text{eng}}$ und $\text{physischen}_{\text{weit}}$ Eigenschaften betrifft. Vor dem Hintergrund dieser Unterscheidung lassen sich zwei unterschiedliche Thesen der Nicht-Identität formulieren:

$\textbf{Nicht-Identität}_{\textbf{weit}}$: Mentale Eigenschaften sind nicht mit $\text{physischen}_{\text{weit}}$ Eigenschaften identisch.
$\textbf{Nicht-Identität}_{\textbf{eng}}$: Mentale und spezialwissenschaftliche Eigenschaften sind nicht mit $\text{physischen}_{\text{eng}}$ Eigenschaften identisch.

Beide Thesen werden im Kontext der Kritik und Verteidigung des nichtreduktiven Physikalismus ausführlich diskutiert. Ausschlaggebend für den nichtreduktiven Physikalismus ist nach meiner Auffassung jedoch ausschließlich ($\text{Nicht-Identität}_{\text{eng}}$). Dies lässt sich durch die folgende Überlegung rechtfertigen:

Erstens sind insbesondere spezialwissenschaftliche Eigenschaften von vornherein trivialerweise mit $\text{physischen}_{\text{weit}}$ Eigenschaften identisch, da $\text{physische}_{\text{weit}}$ Eigenschaften gerade mit Blick auf die Spezialwissenschaften definiert werden. Nicht-reduktive Physikalist*innen vertreten aber auch in Bezug auf spezialwissenschaftliche Eigenschaften eine These der Nicht-Identität. Daher sollte

sich die These der Nicht-Identität, zumindest soweit spezialwissenschaftliche Eigenschaften betroffen sind, auf $\text{physische}_{\text{eng}}$ Eigenschaften beziehen.

Diese Überlegung lässt jedoch offen, dass nicht-reduktive Physikalist*innen *auch* darauf festgelegt sind, dass mentale Eigenschaften nicht mit neurologischen, biologischen oder anderen spezialwissenschaftlichen Eigenschaften identisch sind. Tatsächlich ist dies eine These, die von nicht-reduktiven Physikalist*innen typischerweise vertreten wird. Nach meinem Verständnis ist die Anerkennung dieser These jedoch keine notwendige Bedingung für den nicht-reduktiven Physikalismus. Ausschlaggebend für den nicht-reduktiven Physikalismus ist vielmehr, dass es einerseits überhaupt höherstufige Eigenschaften gibt, die über eine Relation ontologischer Abhängigkeit in eine physikalistische Ontologie integriert werden und dass andererseits auch mentale Eigenschaften in diesem Sinne höherstufig sind. Dies kann jedoch auch dadurch gegeben sein, dass mentale Eigenschaften mit höherstufigen neuronalen Eigenschaften identisch sind, die von $\text{physischen}_{\text{eng}}$ Eigenschaften ontologisch abhängen.

Eine Position, die mentale Eigenschaften mit höherstufigen neuronalen Eigenschaften identifiziert, die von $\text{physischen}_{\text{eng}}$ Eigenschaften ontologisch abhängig sind, hat nach meinem Verständnis alle für den nicht-reduktiven Physikalismus wesentlichen Merkmale: Nach einer solchen Position gelten mentale Eigenschaften als höherstufige Eigenschaften, die von $\text{physischen}_{\text{eng}}$ Eigenschaften ontologisch abhängig sind. Es ergeben sich für eine solche Position daher auch die in Teil 3 dieser Arbeit diskutierten Probleme mit mentaler Verursachung. Die in der Literatur zur Nicht-Identitätsthese größtenteils diskutierte Frage, ob mentale Eigenschaften mit neuronalen Eigenschaften identisch sind, hat daher aus meiner Sicht für sich genommen nicht die hohe Relevanz, die ihr häufig zugeschrieben wird. Eine Lösung des Problems mentaler Verursachung würde aus einer mental-neuronalen Identitätsthese nur folgen, wenn zudem eine Identitätsthese vertreten wird, der zufolge neuronale Eigenschaften mit $\text{physischen}_{\text{eng}}$ Eigenschaften identisch sind.

Als wie allgemein ist die These der Nicht-Identität zu verstehen? Läuft sie darauf hinaus, dass *keine* mentale oder spezialwissenschaftliche Eigenschaft mit einer $\text{physischen}_{\text{eng}}$ Eigenschaft identisch ist? Oder verlangt sie lediglich, dass es mentale oder spezialwissenschaftliche Eigenschaften *gibt*, die nicht mit $\text{physischen}_{\text{eng}}$ Eigenschaften identisch sind? Auf welche der folgenden beiden Thesen, die die Quantifikationsstruktur der These der Nicht-Identität explizit machen, ist der nicht-reduktive Physikalismus also festgelegt?

(Nicht-Identität – A) Für alle mentalen und spezialwissenschaftlichen Eigenschaften M gilt: Es gibt keine $\text{physische}_{\text{eng}}$ Eigenschaft P, so dass gilt: $M=P$.
(Nicht-Identität – E) Es gibt mentale und spezialwissenschaftliche Eigenschaften M, so dass gilt: Es gibt keine $\text{physische}_{\text{eng}}$ Eigenschaft P, so dass gilt: $M=P$.

Die These (Nicht-Identität – A) wäre zu stark. Der nicht-reduktive Physikalismus ist nicht schon dann falsch, wenn es eine mentale oder spezialwissenschaftliche Eigenschaft gibt, die nicht multipel realisiert ist und mit einer $\text{physischen}_{\text{eng}}$ Eigenschaft identifiziert werden kann. Zugleich ist die These (Nicht-Identität – E) schwächer, als das, was nicht-reduktive Physikalist*innen typischerweise behaupten. Denn (Nicht-Identität – E) ist damit kompatibel, dass mentale und spezialwissenschaftliche Eigenschaften, die nicht mit physischen Eigenschaften identisch sind, eine seltene Ausnahme darstellen. Der nicht-reduktive Physikalismus ist jedoch zumindest damit verbunden, dass die Nicht-Identität mentaler und spezialwissenschaftlicher Eigenschaften mit physischen Eigenschaften *die Regel ist*. Bei dieser vagen Bestimmung kann man es ohne Schaden belassen.

2.4.3 Multiple Realisierung

Das für den nicht-reduktiven Physikalismus typische Argument für die These der Nicht-Identität beruht auf der These der multiplen Realisierung. Grob kann dieses *Argument aus der multiplen Realisierung* wie folgt rekonstruiert werden:

Argument aus der multiplen Realisierung:

(MR1) Mentale und spezialwissenschaftliche Eigenschaften sind auf $\text{physischer}_{\text{eng}}$ Ebene multipel realisiert.
(MR2) Wenn mentale und spezialwissenschaftliche Eigenschaften auf $\text{physischer}_{\text{eng}}$ Ebene multipel realisiert sind, dann sind sie nicht mit $\text{physischen}_{\text{eng}}$ Eigenschaften identisch.
(MRK) Also: Mentale und spezialwissenschaftliche Eigenschaften sind nicht mit $\text{physischen}_{\text{eng}}$ Eigenschaften identisch.[59]

Wie sollte die These der multiplen Realisierung verstanden werden, um das Argument aus der multiplen Realisierung zu plausibilisieren?

[59] Vgl. Abschnitt 1.2. in Bickle (2020).

Die Grundidee der These der multiplen Realisierung illustriert Hilary Putnam in der folgenden, vielzitierten Passage:

> Consider what the brain state theorist has to do to make good his claim. He has to specify a physical-chemical state such that any organism (not just a mammal) is in pain if and only if (a) it possesses a brain of a suitable physical-chemical structure; and (b) its brain is in that physical-chemical state. This means that the physical-chemical state in question must be a possible state of a mammalian brain, a reptilian brain, a mollusc's brain (octopuses are mollusca, and certainly feel pain), etc. At the same time, it must not be a possible (physically possible) state of the brain of any physically possible creature that cannot feel pain. Even if such a state can be found, it must be nomologically certain that it will also be a state of the brain of any extra-terrestrial life that may be found that will be capable of feeling pain before we can even entertain the supposition that it may be pain.[60]

Ausgehend von diesem Zitat lässt sich zunächst eine Unterscheidung zwischen tatsächlicher multipler Realisierung und multipler Realisierbarkeit ziehen. Denn Putnam deutet hier zwei unterschiedliche Beispiele an: Das erste Beispiel betrifft die *tatsächliche* multiple Realisierung von Schmerz in unterschiedlichen Spezies: Schmerz hat in Menschen eine andere neurologische Grundlage als in Oktopussen. Das zweite Beispiel betrifft die (bloß mögliche) multiple Realisier*barkeit* von Schmerz in Menschen und Außerirdischen: Schmerzempfindende Außerirdische mit sehr anderen Nervensystemen sind (nomologisch) möglich.[61] Es werden durch diese Beispiele also unterschiedliche Thesen gestützt. Das erste Beispiel stützt eine These tatsächlicher multipler Realisierung: Mentale Eigenschaften sind auf neurologischer Ebene in der aktualen Welt auf unterschiedliche Weise realisiert. Das zweite Beispiel stützt eine schwächere These multipler Realisierbarkeit: Mentale Eigenschaften *können* auf unterschiedliche Weise realisiert sein. Diese zweite These kann auch dann wahr sein, wenn mentale Eigenschaften nicht tatsächlich multipel realisiert sind.

Das Argument aus der multiplen Realisierung kann sich entsprechend einerseits auf tatsächliche multiple Realisierung und andererseits auf bloße multiple Realisierbarkeit beziehen. Bezieht es sich auf tatsächliche multiple Realisierung,

[60] Putnam (1967: 164).

[61] Putnam (1975: 293) deutet zudem die metaphysische Möglichkeit nicht-physischer Realisierer an: "Now, imagine two possible universes, perhaps "parallel worlds", in the science fiction sense, in one of which people have good old fashioned souls, operating through pineal glands, perhaps, and in the other of which they have complicated brains. And suppose the souls in the soul world are functionally isomorphic to the brains in the brain world. Is there any more sense to attaching importance to this difference than to the difference between copper wires and some other wires in the computer?".

sollte die zentrale Prämisse, dass mentale und spezialwissenschaftliche Eigenschaften tatsächlich multipel realisiert sind, durch empirische Evidenzen gestützt werden. Bezieht sich das Argument hingegen auf bloße multiple Realisier*barkeit*, kann die zentrale Prämisse, dass mentale/spezialwissenschaftliche Eigenschaften multipel realisierbar sind, auch durch Vorstellbarkeitsargumente gestützt werden: Da alternative Realisierungen zumindest vorstellbar sind, sind sie auch möglich.[62] Eine Prämisse der tatsächlichen multiplen Realisierung erfordert also eine andere Art von Begründung und erlaubt eine andere Art von Kritik als eine Prämisse der multiplen Realisierbarkeit.[63]

Zugleich ist jedoch festzustellen, dass sich sowohl aus einer These der multiplen Realisierung als auch aus einer These der multiplen Realisierbarkeit ein Problem für Eigenschaftsidentitätsbehauptungen ergibt: Natürlich kann Schmerz nicht mit dem Realisierer in Menschen identisch sein, wenn schon in der aktualen Welt alternative Realisierer instantiiert sind. Denn dann ist Schmerz mit keinem Realisierer aktual ko-instantiiert. Aber Schmerz kann auch dann nicht identisch mit einem aktualen Realisierer sein, wenn es mögliche alternative Realisierer in anderen Welten gibt. Denn dann ist Schmerz nicht *notwendig* mit einem Realisierer ko-instantiiert. Da Eigenschaftsidentität aber notwendige Ko-Instantiierung erfordert, spricht auch dies schon gegen die These, dass Schmerz mit einem Realisierer identisch ist.

Putnams Beispiele betreffen die Beziehung zwischen mentalen Eigenschaften und neurologischen Eigenschaften. Sie stützen also die These, dass mentale Eigenschaften *auf neurologischer Ebene* multipel realisiert bzw. multipel realisierbar sind. Thesen multipler Realisierung können jedoch auch andere Klassen von Eigenschaften betreffen.

---

[62] Solche Vorstellbarkeitsargumente würden jedoch – anders als in der Philosophie des Geistes weit verbreitete Vorstellbarkeitsargument wie das Zombie-Argument gegen den Physikalismus (vgl. Chalmers (2003)) – nicht die Vorstellbarkeit/Möglichkeit von Situationen in Anspruch nehmen, in denen physische Duplikate aktualer denkender und fühlender Personen nicht denken und fühlen. Stattdessen ginge es anders herum um die Vorstellbarkeit/Möglichkeit von Situationen, in denen mentale/spezialwissenschaftliche Duplikate aktualer Objekte andere oder keine physischen Eigenschaften haben.

[63] Der Fokus unterschiedlicher Diskussionen multipler Realisierung unterscheidet sich dabei erheblich. Während z. B. Polger und Shapiro (2016) ausführlich eine These der tatsächlichen multiplen Realisierung kritisieren und die These der bloßen multipler Realisierbarkeit als irrelevant erachten (vgl. auch Baysan (2018: 179) für eine Kritik hieran), hält Stoljar (2010: 120) die Frage nach der tatsächlichen multiplen Realisierung für vernachlässigbar und konzentriert sich auf multiple Realisierbarkeit.

Anknüpfend an ein Beispiel von Jerry Fodor[64] findet sich etwa häufig die These, dass die ökonomische Eigenschaft, ein Austausch von Zahlungsmittel zu sein, auf $\text{physischer}_{\text{eng}}$ Ebene multipel realisiert ist: Sowohl die Übergabe von Bargeld, als auch das Schreiben eines Schecks, die Eingabe einer Pin in ein Kreditkartenleser, oder das Halten eines Smartphones an ein Lesegerät haben aus ökonomischer Sicht die Gemeinsamkeit, ein Austausch von Zahlungsmitteln zu sein. Es ist jedoch wenig plausibel, dass all diese Vorgänge eine relevante $\text{physische}_{\text{eng}}$ Gemeinsamkeit haben. Dies legt die These nahe, dass ökonomische Eigenschaften auf $\text{physischer}_{\text{eng}}$ Ebene multipel realisiert sind.

Nach diesem Schema können zahlreiche andere Thesen multipler Realisierung formuliert und verteidigt werden, die zum Beispiel auch biologische oder chemische Eigenschaften betreffen.[65] Die Thesen setzen dabei immer Eigenschaften, die einer Wissenschaft oder einem Begriffssystem zuzuordnen sind, zu Eigenschaften in Beziehung, die einer *anderen* Wissenschaft oder einem *anderen* Begriffssystem zuzuordnen sind. In diesem Sinne ist multiple Realisierung/Realisierbarkeit eine stufenübergreifende Relation.[66]

Für den nicht-reduktiven Physikalismus ist – unter der Voraussetzung, dass sich die These der Nicht-Identität auf $\text{physische}_{\text{eng}}$ Eigenschaften bezieht – in erster Linie eine These multipler Realisierung oder Realisierbarkeit ausschlaggebend, die die Beziehung zwischen mentalen/spezialwissenschaftlichen Eigenschaften und $\text{physischen}_{\text{eng}}$ Eigenschaften betrifft. Dementsprechend ist die Frage, ob mentale Eigenschaften auf neuronaler Ebene multipel realisiert sind, nach meinem Verständnis nur von abgeleiteter Relevanz für den nicht-reduktiven Physikalismus: Wenn mentale Eigenschaften auf neuronaler Ebene multipel realisiert sind, dann ist zu erwarten, dass sie auch auf $\text{physischer}_{\text{eng}}$ Ebene multipel realisiert sind. Mental-neuronale multiple Realisierung ist also mindestens starke Evidenz für $\text{mental-physische}_{\text{eng}}$ multiple Realisierung.[67] Wenn sie jedoch nicht auf neuronaler Ebene multipel realisiert sind, folgt daraus keinesfalls direkt, dass sie auch auf $\text{physischer}_{\text{eng}}$ Ebene nicht multipel realisiert sind. Für eine solche Folgerung müsste man zusätzlich annehmen, dass die relevanten neuronalen Eigenschaften nicht auf $\text{physischer}_{\text{eng}}$ Ebene multipel realisiert sind.

---

[64] Vgl. Fodor (1974: 103–104).

[65] Vgl. z. B. Fang (2018) für eine Verteidigung multipler Realisierung in der Biologie.

[66] Zum Verhältnis zwischen wissenschaftlichen Disziplinen und der Rede von Stufen siehe auch Abschnitt 2.5.5.

[67] Eine logische Folgerung von mental-neuronaler auf $\text{mental-physische}_{\text{eng}}$ multiple Realisierung ist hingegen nach meiner Einschätzung nicht gültig. Es wäre ja durchaus möglich, dass sich Gemeinsamkeiten zwischen Schmerz-instantiierenden Objekten auf neuronaler Ebene nicht finden, aber auf $\text{physischer}_{\text{eng}}$ Ebene wieder auftauchen.

Nicht-reduktive Physikalist*innen sollten dabei die These vertreten, dass *die meisten oder zumindest viele* mentale und spezialwissenschaftliche Eigenschaften auf $\text{physischer}_{\text{eng}}$ Ebene multipel realisiert sind.[68] Eine solche allgemeine These multipler Realisierung beruht auf einer großen Anzahl von spezifischeren Thesen. Diese spezifischeren Thesen sind durch die folgende Form ausgezeichnet:

> **(SMR)** Eigenschaft **F**, die der Klasse der mentalen oder der Klasse der spezialwissenschaftlichen Eigenschaft angehört, ist multipel realisiert/realisierbar durch die $\text{physischen}_{\text{eng}}$ Eigenschaften $\mathbf{R_1}$ _ $\mathbf{R_n}$.

Wie ist eine solche These multipler Realisierung nun genauer zu verstehen? Wie ist insbesondere das Verhältnis zwischen der realisierten Eigenschaft **F** und den realisierenden Eigenschaften $\mathbf{R_1 - R_n}$ genauer zu verstehen?

Naheliegend ist es, die Beziehung zwischen den einzelnen möglichen Realisierern von **F** und **F** als die Beziehung der Realisierung aufzufassen.[69] Wenn **F** durch **R1 – Rn** multipel realisiert ist, dann ist **F** durch **R1** realisiert, durch **R2** realisiert usw. Es gibt jedoch gute Gründe, die These der multiplen Realisierung nicht mit einer solchen These der Realisierung zu verknüpfen:

Erstens ist die Realisierungsrelation nur eine Option, die für den nicht-reduktiven Physikalismus charakteristische These der ontologischen Abhängigkeit zu explizieren. Grounding und Determination sind zwei Alternativen.[70] Auch nicht-reduktive Physikalist*innen, die sich nicht darauf festlegen, dass mentale Eigenschaften durch $\text{physische}_{\text{eng}}$ Eigenschaften *realisiert* sind, berufen sich auf die These der multiplen Realisierung. Die These der multiplen Realisierung sollte also neutral gegenüber unterschiedlichen Explikationen der Relation der ontologischen Abhängigkeit sein. Dies legt nahe, die Relation zwischen **F** und $\mathbf{R_1 - R_n}$ neutral als die Relation der ontologischen Abhängigkeit zu bestimmen.[71]

Zweitens ist jedoch – wie Thomas Polger[72] betont – auch diese Option nicht unproblematisch: Denn die Relation der ontologischen Abhängigkeit ist plausiblerweise eine asymmetrische Relation: Wenn **F** von $\mathbf{R_1}$ ontologisch abhängig ist, dann ist $\mathbf{R_1}$ nicht von **F** ontologisch abhängig. Daher schließt schon die

---

[68] Siehe auch die entsprechende Klärung der Allgemeinheit der These der Nicht-Identität in Abschnitt 2.4.2.

[69] Auf die Beziehung der Realisierung gehe ich in Abschnitt 2.5.3. noch ein.

[70] Vgl. die Einführung dieser Begriffe in den Abschnitten 2.5.4. und 2.5.6.

[71] Ich spreche im Folgenden dennoch in einem ungenauen Sinne von ‚Realisierern', wenn ich R1-Rn anspreche. Hiermit soll aber eben nicht verbunden sein, dass F zu R1-Rn in der Relation der Realisierung steht, die in 2.5.3. thematisiert wird.

[72] Polger (2015: 865–866).

These, dass **F** von $\mathbf{R_1}$ ontologisch abhängig ist, aus, dass **F** mit $\mathbf{R_1}$ identisch ist. Eine Eigenschaft kann nicht in einer asymmetrischen Relation zu sich selbst stehen. Die These der Nicht-Identität würde also aus der isolierten Betrachtung der Beziehung zwischen *einem* ‚Realisierer' und **F** folgen. In der Konsequenz wäre ein Verweis auf *multiple* Realisierung im Kontext des Arguments aus der multiplen Realisierung überflüssig. Eine angemessene Interpretation des Arguments aus der multiplen Realisierung sollte hingegen dem Umstand Rechnung tragen, dass erst die *multiple* Realisierung (und nicht schon die bloße ‚Realisierung') von **F** gegen eine Eigenschaftsidentifikation spricht.

Ein Ausweg besteht darin, die Beziehung zwischen **F** und $\mathbf{R_1} - \mathbf{R_n}$ in rein modalen Begrifflichkeiten zu bestimmen. Gehen wir vereinfachend davon aus, dass **F** nur zwei mögliche Realisierer $\mathbf{R_1}$ und $\mathbf{R_2}$ hat. Die Beziehung zwischen **F** und $\mathbf{R_1}$ bzw. $\mathbf{R_2}$ kann dann durch die folgenden modalen Annahmen charakterisiert werden:

(MA1) Notwendig gilt: Wenn $\mathbf{R_1}$ instantiiert ist, dann ist **F** instantiiert.
(MA2) Notwendig gilt: Wenn $\mathbf{R_2}$ instantiiert ist, dann ist **F** instantiiert.[73]

Diese Annahmen sind neutral gegenüber der Frage, wie ontologische Abhängigkeit expliziert werden sollte. Auch folgt aus diesen Annahmen für sich genommen noch nicht, dass **M** nicht mit $\mathbf{R_1}$ oder $\mathbf{R_2}$ identisch ist: Die These, dass **F** nicht mit $\mathbf{R_1}$ identisch ist, ergibt sich nicht aus einer isolierten Betrachtung von **F** und $\mathbf{R_1}$.

Stattdessen ist es ausschlaggebend für die Konklusion, dass **F** nicht mit $\mathbf{R_1}$ identisch ist, dass es *andere Realisierer von* **F** gibt. Denn erst dies etabliert, dass keiner der Realisierer immer dann instantiiert ist, wenn **F** instantiiert ist. Wenn Schmerz beispielsweise in Oktopussen anders realisiert ist als in Menschen, dann ist Schmerz gelegentlich instantiiert, ohne dass der menschliche Realisierer von Schmerz instantiiert ist. Da Eigenschaftsidentität notwendige Ko-Instantiierung voraussetzt, kann Schmerz also nicht mit dem menschlichen Realisierer identisch sein.

Aus der These, dass **F** tatsächlich durch $\mathbf{R_1}$ und $\mathbf{R_2}$ multipel realisiert ist, ergibt sich also der folgende Zusammenhang:

---

[73] Ich setze hier voraus, dass **R1** und **R2** totale Realisierer von **F** sind und nicht bloß Kern-Realisierer (vgl. Shoemaker (1981)). Diese Begriffe werden in 2.5.3 eingeführt. Wenn **R1** und **R2** bloße Kernrealisierer wären, würden sie **F** nicht necessitieren. Dann aber wäre wiederum eine Identifikation schon bei der isolierten Betrachtung von **R1** und **F** ausgeschlossen und der Verweis auf *multiple* Realisierung wäre überflüssig.

(MA3) Es gilt nicht: Wenn **F** instantiiert ist, dann ist $\mathbf{R_1}$ instantiiert.
(MA4) Es gilt nicht: Wenn **F** instantiiert ist, dann ist $\mathbf{R_2}$ instantiiert.

Aus der These, dass **F** durch **R1** und **R2** multipel realisier*bar* ist, ergibt sich hingegen der folgende Zusammenhang:

(MA5) Es gilt nicht *notwendig*: Wenn **F** instantiiert ist, dann ist $\mathbf{R_1}$ instantiiert.
(MA6) Es gilt nicht *notwendig*: Wenn **F** instantiiert ist, dann ist $\mathbf{R_2}$ instantiiert.

Man sieht hier noch einmal, dass tatsächliche multiple Realisierung eine stärkere These ist als multiple Realisierbarkeit: Denn (MA3) impliziert (MA5), ohne von (MA5) impliziert zu werden. Zugleich wird noch einmal deutlich, dass schon multiple Realisierbarkeit einen Schluss auf Thesen der Nicht-Identität zulässt: Denn auch aus (MA5) folgt bereits, dass **F** nicht mit $\mathbf{R_1}$ identisch ist.

Aus (MA5) und (MA6) folgt jedoch noch nicht, dass es keine $\text{physische}_{\text{eng}}$ Eigenschaft gibt, mit der **F** identisch ist. Dass **F** nicht mit ihren Realisierern notwendig ko-instantiiert ist, heißt nicht zwangsläufig, dass **F** mit keiner $\text{physischen}_{\text{eng}}$ Eigenschaft notwendig ko-instantiiert ist.[74]

Gehen wir jedoch davon aus, dass jede $\text{physische}_{\text{eng}}$ Eigenschaft *R*, für die notwendig gilt: Wenn *R* instantiiert ist, dann ist **F** instantiiert, im relevanten Sinne als Realisierer von **F** zählt, löst sich dieses Problem auf: $\text{Physische}_{\text{eng}}$ Eigenschaften, die keine Realisierer von **F** sind, kommen dann nicht für eine Identität mit **F** in Frage, weil sie **F** nicht necessitieren. Die Realisierer von **F** kommen hingegen nicht in Frage, weil sie nicht von **F** necessitiert werden. Unter dieser Voraussetzung folgt aus der oben formulierten These multipler Realisierung (SMR) also weiterhin:

(MA7) Es gibt keine $\text{physische}_{\text{eng}}$ Eigenschaft *P*, so dass notwendig gilt: **F** ist genau dann instantiiert, wenn *P* instantiiert ist.

Diese These hat nun die unmittelbare Konsequenz, dass **F** mit *keiner* $\text{physischen}_{\text{eng}}$ Eigenschaft identisch ist.

Kommen wir noch kurz zu Einwänden gegen die These der multiplen Realisierung: Thesen multipler Realisierung beruhen – wie deutlich geworden sein sollte – darauf, dass mehrere Objekte dieselbe mentale bzw. spezialwissenschaftliche Eigenschaft instantiieren, ohne aber eine entsprechende physische

---

[74] Vgl. auch hierzu Polger (2015: 866).

Eigenschaft zu teilen. Sie enthalten somit sowohl eine Gleichheits- (oder zumindest Ähnlichkeits-)these als auch eine Ungleichheits- (oder Unähnlichkeits-) these: Manche Objekte sind in einer mentalen oder spezialwissenschaftlichen Hinsicht gleich oder ähnlich, ohne in einer physischen Hinsicht gleich oder ähnlich zu sein. Beide Elemente von Thesen der multiplen Realisierung können angegriffen werden: Es kann erstens bezweifelt werden, dass eine mentale bzw. spezialwissenschaftliche Gleichheit oder Ähnlichkeit vorliegt und es kann zweitens bezweifelt werden, dass es keine zugehörige physische Gleichheit oder Ähnlichkeit gibt.[75]

Beide Strategien lassen sich am besten anhand von spezifischen Thesen multipler Realisierung veranschaulichen. Nehmen wir in Anlehnung an das Putnam-Zitat die folgende These der multiplen Realisierung als Ausgangspunkt:

> **Multiple Realisierung – Putnams Beispiel**: Die mentale Eigenschaft, Schmerz zu haben, ist durch die neurologischen Eigenschaften $\mathbf{N_1}$ und $\mathbf{N_2}$ multipel realisiert – wobei $\mathbf{N_1}$ eine neuronale Eigenschaft ist, die bei Menschen aber nicht bei Oktopussen zu finden ist, während $\mathbf{N_2}$ eine neuronale Eigenschaft ist, die bei Oktopussen aber nicht bei Menschen zu finden ist.

Um diese These zu widerlegen, kann man entweder bestreiten, dass Menschen und Oktopusse die mentale Eigenschaft teilen, Schmerzen zu haben. Oder man kann bestreiten, dass Menschen und Oktopusse, die Schmerzen haben, keine neuronale Eigenschaft teilen, die notwendig mit Schmerzen ko-instantiiert ist.

Die erste Strategie führt zu einer Aufteilung verschiedener mentaler Eigenschaften, die aber jeweils mit einer neuronalen Eigenschaft identifiziert werden können: Es gibt Schmerz-in-Menschen und Schmerz-in-Oktopussen. Schmerz-in-Menschen ist notwendig ko-instantiiert und identisch mit $\mathbf{N_1}$, während Schmerz-in-Oktopussen notwendig ko-instantiiert und identisch mit $\mathbf{N_2}$ ist. Eine unqualifizierte Schmerz-Eigenschaft, die Menschen und Oktopusse teilen, gibt es nach dieser Ansicht nicht. Tatsächlich besteht also auf mentaler Ebene ebensowenig eine Gleichheit oder Ähnlichkeit wie auf physischer Ebene.[76]

Die Plausibilität dieser Strategie ist zu einem guten Stück davon abhängig, wie weit die multiple Realisierbarkeit mentaler Eigenschaften geht: Gibt es zum Beispiel bloß zwei mögliche Realisierer von Schmerz – einen in Oktopussen und einen in Menschen? Oder gibt es für jede Spezies einen spezifischen Realisierer? Bleibt es bei speziesspezifischen Realisierungen, oder ist Schmerz auch bei

[75] Vgl. Funkhouser (2007a: 305).

[76] Vgl. z. B. Kim (1992: 19–21).

unterschiedlichen Menschen unterschiedlich realisiert? Kann Schmerz schließlich sogar beim selben Menschen zu verschiedenen Zeitpunkten auf verschiedene Weise realisiert sein?[77]

Wenn eine mentale Eigenschaft wie Schmerz etwa genau zwei mögliche Realisierer in Kraken und in Menschen hat, mag eine Aufspaltung der mentalen Eigenschaft in Kraken-Schmerz und in Menschen-Schmerz nachvollziehbar sein. Wenn Schmerz hingegen in jedem Lebewesen und sogar innerhalb eines bestimmten Lebewesens zahlreiche Realisierer hat, ist eine entsprechende Aufspaltung auf mentaler Ebene eher unplausibel und ad hoc. Die Strategie wäre bei einer relativ weitgehenden multiplen Realisierung gewöhnlicher mentaler Eigenschaften schnell darauf festgelegt, eine anti-realistische Haltung gegenüber allen gewöhnlichen mentalen Eigenschaften einzunehmen: Wenn Schmerz zum Beispiel individuenspezifisch realisiert ist, so dass es bei jedem Menschen einen ihm eigenen Schmerz-Realisierer gibt, würde aus der diskutierten Strategie folgen, dass es keine individuenübergreifende Schmerzeigenschaft gibt.

Zudem stellt sich die Frage, ob die Unterscheidungen, die im Rahmen dieser Strategie aufgrund von multipler Realisierung auf mentaler Ebene getroffen werden, aus Sicht der (Alltags-)Psychologie gut motiviert werden können. Gibt es Kriterien für diese Unterscheidungen, die in der Sprache der Psychologie ausgedrückt werden können? Gibt es psychologische Gesetze, die zwischen den verschiedenen Arten von Schmerz zu differenzieren helfen? Oder sind die resultierenden Unterscheidungen sozusagen von außen an die Psychologie herangetragen – so dass der einzige Grund, diese Unterscheidungen in das psychologische Vokabular einzuführen, darin besteht, dass die ‚verschiedenen' psychologischen Arten auf neuronaler Ebene unterschiedlich realisiert sind? Wenn sich die entsprechenden Unterscheidungen nicht aus der Psychologie heraus motivieren lassen, spricht dies gegen diese Strategie. Denn sie würde dazu führen, dass die Psychologie methodologisch nicht autonom ist und die psychologische Einteilung der Welt ontologisch nicht ernstgenommen wird. Auch hier wird es sicherlich umso schwerer, die Unterscheidungen, die sich aus multipler Realisierung ergeben, innerhalb der Psychologie zu rechtfertigen, je stärker Schmerz multipel realisiert ist.

Die zweite Strategie, die These multipler Realisierung im obigen Beispiel zu bestreiten, besteht darin, sich auf die Suche nach relevanten neurologischen Gemeinsamkeiten zwischen Oktopussen und Menschen zu machen. So könnte man etwa argumentieren, dass es grob charakterisierte neurologische Eigenschaften gibt, die Menschen mit Schmerzen und Oktopusse mit Schmerzen

---

[77] Vgl. Bickle (2020).

teilen.[78] Dass Menschen mit Schmerzen die neuronale Eigenschaft $\mathbf{N_1}$ haben und Oktopusse mit Schmerzen die neuronale Eigenschaft $\mathbf{N_2}$ haben, schließt schließlich nicht aus, dass sowohl Menschen mit Schmerzen als auch Oktopusse mit Schmerzen die neuronale Eigenschaft **N** haben und **N** speziesübergreifend notwendig mit Schmerz ko-instantiiert ist. Gestützt werden kann eine solche Hoffnung auf neurologische Gemeinsamkeiten, die mit psychologischen Gemeinsamkeiten einhergehen, durch die Beobachtung, dass die Begriffsbildung in den Neurowissenschaften sich auch an psychologischen Ähnlichkeiten orientiert. Eine angenommene psychologische Ähnlichkeit zwischen z. B. verschiedenen Spezies motiviert daher eine gemeinsame neurologische Klassifikation. Diese Methodologie schränkt die Wahrscheinlichkeit multipler Realisierung mentaler Eigenschaften auf neurologischer Ebene von vornherein ein.[79]

Diese Überlegung verdeutlicht, dass die multiple Realisierung mentaler Eigenschaften auf neurologischer Ebene nicht zu schnell als etabliert gelten sollte. Es ist sehr einfach, den Anschein einer multiplen Realisierung hervorzurufen, wenn man sehr grob charakterisierte mentale Eigenschaften zu sehr fein charakterisierten neuronalen Eigenschaften in Beziehung setzt. Wichtig ist jedoch, dass eine interessante multiple Realisierung mentaler Eigenschaften auf neurologischer Ebene nicht bloß voraussetzt, dass sich neurologische *Unterschiede* zwischen den Objekten, die die mentalen Eigenschaften instantiieren, finden lassen. Stattdessen fängt interessante multiple Realisierung dort an, wo sich *keine* (einzigartigen) neurologischen *Gemeinsamkeiten* zwischen den Objekten, die die mentale Eigenschaft instantiieren, finden lassen. Es reicht also in unserem Beispiel nicht, dass sich eine neurologische Unterscheidung zwischen $\mathbf{N_1}$ (bei Menschen) und $\mathbf{N_2}$ (bei Oktopussen) ziehen lässt. Es muss zudem etabliert sein, dass sich $\mathbf{N_1}$ und $\mathbf{N_2}$ nicht als Spezifizierungen einer allgemeineren neurologischen Eigenschaft **N** verstehen lassen, die mit Schmerz notwendig ko-instantiiert ist.

Die Frage nach der multiplen Realisierung mentaler Eigenschaften auf neuronaler Ebene ist daher im Einzelfall auf der Grundlage empirischer Studien zu prüfen. Die Konsequenzen aus diesem Punkt für den nicht-reduktiven Physikalismus sind jedoch aus meiner Sicht ohnehin nicht sehr weitreichend. Selbst wenn mentale Eigenschaften nicht *auf neuronaler Ebene* multipel realisiert sind, scheint es noch immer sehr wahrscheinlich, dass sie auf niedrigeren Ebenen multipel realisiert sind. Betrachtet man etwa direkt das Verhältnis zwischen mentalen

[78] Vgl. Bechtel und Mundale (1999).

[79] Vgl. Bechtel und Mundale (1999: 190): “Identifying brain areas through neuroimaging depends critically on the cognitive tasks subjects are asked to perform; thus, the possibility of multiple realizability is restricted at the outset.”

Eigenschaften und $\text{physischen}_{\text{eng}}$ Eigenschaften, scheint eine saubere eins-zu-eins Korrelation völlig aussichtslos. Welche $\text{physische}_{\text{eng}}$ Eigenschaft könnte notwendig ko-instantiiert sein mit der mentalen Eigenschaft, Schmerz zu haben? Es gibt hier nicht einmal entfernt aussichtsreiche Kandidaten.

Die besprochenen Einwände gegen die multiple Realisierung mentaler Eigenschaften auf neuronaler Ebene nehmen der für den nicht-reduktiven Physikalismus ausschlaggebenden These entsprechend nichts an ihrer Plausibilität: Auf der Ebene der $\text{physischen}_{\text{eng}}$ Eigenschaften sind sowohl mentale als auch spezialwissenschaftliche Eigenschaften multipel realisiert.[80] Zudem scheint die multiple Realisierung mentaler Eigenschaften durch $\text{physische}_{\text{eng}}$ Eigenschaften weitreichend zu sein: Es wirkt nicht sehr wahrscheinlich, dass sich recht große – zum Beispiel speziesspezifische – Klassen von Realisierern auf Ebene der Physik spezifizieren lassen. Stattdessen ist eine multiple Realisierung auch unter verschiedenen Menschen und unter individuellen Menschen zu verschiedenen Zeitpunkten zu erwarten.

Schon diese These der multiplen Realisierung mentaler und spezialwissenschaftlicher Eigenschaften durch $\text{physische}_{\text{eng}}$ Eigenschaften liefert eine gute Motivation für den nicht-reduktiven Physikalismus: Wenn mentale und spezialwissenschaftliche Eigenschaften durch $\text{physische}_{\text{eng}}$ Eigenschaften multipel realisiert sind, kann man sie nicht mit $\text{physischen}_{\text{eng}}$ Eigenschaften identifizieren. Wenn man sie nicht mit $\text{physischen}_{\text{eng}}$ Eigenschaften identifizieren kann, muss man sie als Physikalist*in entweder eliminieren oder den Weg der Abhängigkeit wählen. Da eine Elimination nicht sonderlich attraktiv wirkt, bleibt also der Weg der Abhängigkeit und somit der nicht-reduktive Physikalismus.[81]

Ein letzter Einwand gegen die These multipler Realisierung und die dadurch motivierte These der Nicht-Identität betrifft die Möglichkeit, disjunktive physische Eigenschaften zu konstruieren, die mit mentalen Eigenschaften notwendig ko-instantiiert sind: Selbst wenn es keine 'gewöhnliche' $\text{physische}_{\text{eng}}$ Eigenschaft gibt, die mit Schmerz notwendig ko-instantiiert ist, gibt es noch immer eine (möglicherweise recht große) Zahl an möglichen $\text{physischen}_{\text{eng}}$ Realisierern von Schmerz. Nun – so die Idee – lässt sich unter Bezug auf die $\text{physischen}_{\text{eng}}$ Realisierer der mentalen Eigenschaft einfach eine disjunktive $\text{physische}_{\text{eng}}$ Eigenschaft definieren, die mit der mentalen Eigenschaft notwendig ko-instantiiert ist. Wenn beispielsweise die physischen Eigenschaften $\mathbf{P_1}$ und $\mathbf{P_2}$ die einzigen möglichen Realisierer von **M** sind, wird **M** notwendig mit der $\text{physischen}_{\text{eng}}$

[80] Vgl. auch Funkhouser (2007b: 475–476).

[81] Die hier zugrunde gelegte Unterscheidung zwischen Identifikation, Elimination und Weg der Abhängigkeit habe ich in Abschnitt 2.1.2 eingeführt.

Eigenschaft $\mathbf{P_1} \lor \mathbf{P_2}$ ko-instantiiert sein. Es besteht somit die Möglichkeit, **M** mit der disjunktiven $\text{physischen}_{\text{eng}}$ Eigenschaft $\mathbf{P_1} \lor \mathbf{P_2}$ zu identifizieren.

Die naheliegende und korrekte Erwiderung auf diesen Einwand besteht in dem Hinweis darauf, dass spärliche Eigenschaften nicht unter Disjunktion geschlossen sind. Daraus, dass **F** eine spärliche Eigenschaft ist und dass **G** eine spärliche Eigenschaft ist, folgt nicht, dass $\mathbf{F} \lor \mathbf{G}$ eine spärliche Eigenschaft ist. Rot zu sein ist eine spärliche Eigenschaft: Die roten Dinge in der Welt verbindet eine echte Gemeinsamkeit. Rund zu sein ist ebenfalls eine spärliche Eigenschaft: Die runden Dinge in der Welt verbindet eine echte Gemeinsamkeit. Aber rot oder rund zu sein ist keine spärliche Eigenschaft. Die Dinge, die entweder rot oder rund sind, sind nicht durch eine echte Gemeinsamkeit in der Welt verbunden. Spärliche $\text{physische}_{\text{eng}}$ Eigenschaften lassen sich also nicht einfach durch Definition in die Welt bringen. Entsprechend etabliert die disjunktive Strategie keineswegs, dass es eine spärliche $\text{physische}_{\text{eng}}$ Eigenschaft gibt, die mit einer mentalen Eigenschaft notwendig ko-instantiiert ist.

Wer also die multiple Realisierung mentaler Eigenschaften akzeptiert und mentale Eigenschaften mit disjunktiven $\text{physischen}_{\text{eng}}$ Eigenschaften identifiziert, gibt letztlich den ontologischen Realismus in Bezug auf mentale Eigenschaften auf. Wenn mentale Eigenschaften mit disjunktiven $\text{physischen}_{\text{eng}}$ Eigenschaften identisch sind, und die disjunktiven $\text{physischen}_{\text{eng}}$ Eigenschaften keine spärlichen Eigenschaften sind, können auch die entsprechenden mentalen Eigenschaften keine spärlichen Eigenschaften sein. Die disjunktive Strategie ist also nicht mit einem ontologischen Realismus in Bezug auf mentale Eigenschaften kompatibel.

Ein weiteres Problem für die disjunktive Strategie hat mit mentaler Verursachung zu tun. Denn die wichtigste Motivation dafür, mentale Eigenschaften überhaupt mit physischen Eigenschaften zu identifizieren, besteht in der damit zusammenhängenden Lösung des Problems mentaler Verursachung: Da $\text{physische}_{\text{eng}}$ Eigenschaften unbestritten physische Wirkungen haben können, führt eine Identifikation mentaler Eigenschaften mit $\text{physischen}_{\text{eng}}$ Eigenschaften direkt dazu, dass auch mentale Eigenschaften $\text{physische}_{\text{eng}}$ Wirkungen haben können. Diese vermeintlich elegante und einfache Lösung des Problems mentaler Verursachung gilt als der Hauptvorzug des reduktiven Physikalismus.

Diese Lösung ist unter Verwendung der disjunktiven Strategie jedoch problematisch: Es ist keinesfalls klar, dass disjunktive Eigenschaften (oder genauer: die zugehörigen Ereignisse) überhaupt als Ursachen in Frage kommen. Es kommt so der Verdacht auf, dass die disjunktive Strategie mentale Eigenschaften mit den *falschen* $\text{physischen}_{\text{eng}}$ Eigenschaften identifiziert. Wenn die physischen Eigenschaften, mit denen mentale Eigenschaften im Rahmen der disjunktiven Strategie

identifiziert werden, selbst epiphänomenal sind, kann die disjunktive Strategie die kausale Effektivität mentaler Eigenschaften natürlich nicht retten.[82]

Und selbst wenn nichts Grundsätzliches dagegen spricht, disjunktive Eigenschaften/Ereignisse als Ursachen zu klassifizieren, stellt sich noch immer die Frage, weshalb disjunktive $\text{physische}_{\text{eng}}$ Eigenschaften nicht ebenso in kausale Konkurrenz zu nicht-disjunktiven $\text{physischen}_{\text{eng}}$ Eigenschaften treten sollten, wie mentale Eigenschaften im nicht-reduktiven Physikalismus: Wenn schon **P1** alle kausale Arbeit leistet, was soll die disjunktive Eigenschaft **P1** ∨ **P2** dann noch tun?[83]

Ein drittes Problem für die disjunktive Strategie ist das folgende: Nimmt man die (u. a. von Putnam ins Spiel gebrachte)[84] Möglichkeit nicht-physischer Realisierer mentaler Eigenschaften ernst, wird es keine *rein physischen* disjunktiven Eigenschaften geben, die mit einer mentalen Eigenschaft notwendig ko-instantiiert sind. Wenn es Welten gibt, in denen Schmerz durch nicht-physische Eigenschaften einer Seele realisiert ist, ist erst eine disjunktive Eigenschaft, die auch diesen nicht-physischen Realisierer als Disjunkt enthält, mit Schmerz notwendig ko-instantiiert. Schmerz kann dann nicht mit einer *physischen* disjunktiven Eigenschaft identisch sein, sondern nur mit einer disjunktiven Eigenschaft, die auch nicht-physische Eigenschaften als Disjunkte enthält. In diesem Falle könnte man nicht davon sprechen, dass die disjunktive Strategie einen reduktiven Physikalismus etabliert.[85] Die disjunktive Strategie ist also auf eine Ablehnung der *Möglichkeit* nicht-physischer Realisierer festgelegt. Insofern man den Physikalismus aber als *kontingente* These vertritt, ist diese Ablehnung der Möglichkeit nicht-physischer Realisierer nicht sehr plausibel.

2.5 Ontologische Abhängigkeit

Die nicht-reduktiven Kernthesen des nicht-reduktiven Physikalismus etablieren, dass mentale und spezialwissenschaftliche Eigenschaften existieren und nicht mit physischen Eigenschaften identisch sind. Trotz dieser nicht-reduktiven Thesen versteht sich der nicht-reduktive Physikalismus als physikalistische Position. Da der Physikalismus (wie ich in Abschnitt 2.1.2. verdeutlicht habe) durch

[82] Vgl. hierzu z. B. Lewis (1994: 420–21) sowie Horgan (2001: 207). Für eine allgemeine Diskussion disjunktiver Ursachen siehe Sartorio (2006).

[83] Mehr zu dem entsprechenden Problem für den nicht-reduktiven Physikalismus in Teil 3.

[84] Vgl. Putnam (1975: 293).

[85] Vgl. auch Francescotti (2014: 18–22).

die These ausgezeichnet ist, dass alle Eigenschaften letztlich physisch sind, steht der nicht-reduktive Physikalismus vor der folgenden Herausforderung: Es muss verdeutlicht werden, wie mentale und spezialwissenschaftliche Eigenschaften *letztlich* physisch sein können, auch wenn sie nicht im Wortsinne physisch sind: Wie können mentale und spezialwissenschaftliche Eigenschaften letztlich physisch sein, obgleich sie nicht mit physischen Eigenschaften identisch sind?

Hier hat nun die These der ontologischen Abhängigkeit ihren Auftritt. In Abschnitt 2.1.4. habe ich diese These vorläufig wie folgt formuliert:

> **(OA) Ontologische Abhängigkeit**: Mentale und spezialwissenschaftliche Eigenschaften sind von physischen Eigenschaften in einem hinreichend starken Sinne ontologisch abhängig.

Die Grundidee ist hier, dass die Existenz mentaler und spezialwissenschaftlicher Eigenschaften, die nicht mit physischen Eigenschaften identisch sind, aus physikalistischer Perspektive akzeptabel ist, insofern diese Eigenschaften in einem hinreichend starken Sinne von physischen Eigenschaften ontologisch abhängig sind. Mentale und spezialwissenschaftliche Eigenschaften werden im nicht-reduktiven Physikalismus also nicht über den Weg der Identifikation, sondern über den Weg der Abhängigkeit in die physikalistische Metaphysik integriert. Diese Strategie wirft einige Fragen auf, auf die ich im Folgenden eingehen möchte.

Dabei verfahre ich wie folgt: In Abschnitt 2.5.1. spezifiziere ich zunächst einige allgemeine Anforderungen, die die Relation der ontologischen Abhängigkeit im Rahmen des nicht-reduktiven Physikalismus erfüllen sollte. Die Abschnitte 2.5.2 bis 2.5.6. diskutieren knapp einige Kandidaten für die gesuchte Relation der ontologischen Abhängigkeit: Supervenienz, Realisierung, Determination, Mikrobasiertheit und Grounding. In Abschnitt 2.5.7. ziehe ich eine knappe Zwischenbilanz. Die verbleibenden Abschnitte 2.5.8. und 2.5.9. diskutieren die Beziehung zwischen ontologischer Abhängigkeit und zwei weiteren Begriffen, die häufig mit dem nicht-reduktiven Physikalismus in Verbindung gebracht werden: Abschnitt 2.5.8. behandelt die Beziehung zwischen ontologischer Abhängigkeit und absoluter Fundamentalität, Abschnitt 2.5.9. die Beziehung zwischen ontologischer Abhängigkeit und einem Stufenmodell der Realität.

2.5.1 Ontologische Abhängigkeit und ontologische Unschuld

Die Grundidee des Physikalismus ist es, dass alle Eigenschaften letztlich nicht mehr als (‚nothing over and above') physische Eigenschaften sind. Die problematischen Eigenschaften – wie mentale oder spezialwissenschaftliche Eigenschaften – sind in einem gewissen Sinne ontologisch unschuldig: Sie führen nichts Zusätzliches in die Ontologie ein.[86]

Damit der nicht-reduktive Physikalismus Aussicht auf Erfolg haben kann, müssen solcherlei Thesen der ontologischen Unschuld mit der These der Nicht-Identität vereinbar sein: Die problematischen Eigenschaften sind nicht strikt identisch mit physischen Eigenschaften, sie sind jedoch auch ‚nicht mehr als' physische Eigenschaften. Die wesentliche Funktion der Relation der ontologischen Abhängigkeit im Kontext des nicht-reduktiven Physikalismus ist es daher, eben dies zu leisten: Wenn nicht-physische Eigenschaften von physischen Eigenschaften in hinreichend starkem Sinne ontologisch abhängig sind, dann sind sie nicht mehr als physische Eigenschaften und somit *letztlich* physisch.

Was aber kann es heißen, dass eine Eigenschaft zwar nicht strikt identisch mit einer physischen Eigenschaft ist, aber eben auch nicht mehr ist als physische Eigenschaften?

Zunächst soll diese Idee mit einer verbreiteten theologischen Metapher veranschaulicht werden.[87] Wenn alle Eigenschaften nicht mehr als physische Eigenschaften sind, muss man sich die (gedachte) Schöpfung der Welt ungefähr so vorstellen: Gott kümmert sich um die Verteilung der physischen Eigenschaften und Objekte und erlässt physische Naturgesetze. Damit ist die Arbeit getan. Gott muss sich nicht zusätzlich darum kümmern, dass es auch spezialwissenschaftliche und mentale Eigenschaften gibt und dass diese auch so verteilt sind, wie gewünscht. Dies ergibt sich automatisch aus der Verteilung der physischen Eigenschaften. *Alle* Eigenschaften stehen fest, wenn die physischen Eigenschaften feststehen. In diesem Sinne haben die nicht-physischen Eigenschaften keine unabhängige Existenz. Ihre Existenz ergibt sich einfach aus der Existenz der physischen Eigenschaften.

Ein wichtiger Punkt an diesem fiktiven Schöpfungsmythos ist, dass aus dieser Geschichte nicht folgt, dass spezialwissenschaftliche und mentale Eigenschaften identisch mit physischen Eigenschaften sind. Es wird nicht verlangt, dass alle

---

[86] Vgl. hierzu auch Abschnitt 3.4. in McLaughlin und Bennett (2018).

[87] Vgl. z. B. Kripke (1980: 153–154) oder Crane (1991).

Objekte, die beispielsweise glauben, dass Paris in Frankreich ist, notwendigerweise eine bestimmte physische Eigenschaft teilen. Vielmehr können sie ganz unterschiedliche physische Eigenschaften haben. Jedoch ergibt sich die Eigenschaft, zu glauben, dass Paris in Frankreich ist, in jedem Falle aus den physischen Eigenschaften.

Wenn nicht-physische Eigenschaften ‚nicht mehr als' physische Eigenschaften sind, ohne mit physischen Eigenschaften identisch zu sein, sollte es ferner plausibel sein, dass die nicht-physischen Eigenschaften nur deshalb instantiiert sind, *weil* die physischen Eigenschaften instantiiert sind. Die nicht-physischen Eigenschaften kommen *kraft* der physischen Eigenschaften in die Welt. Sie sind *aufgrund von* physischen Eigenschaften instantiiert. All diese Redewendungen verweisen auf eine spezifische Art von *metaphysischer Erklärung*. Die These der ontologischen Abhängigkeit hat also auch die Funktion, solche metaphysischen Erklärungsbehauptungen zu begründen.[88]

Aus dieser Motivation folgen einige Anforderungen an eine geeignete Beziehung ontologischer Abhängigkeit im Rahmen des nicht-reduktiven Physikalismus. Bei der gesuchten Relation sollte es sich erstens um eine *asymmetrische* oder zumindest *nicht-symmetrische* Relation handeln: Schließlich sollen die nicht-physischen Eigenschaften sich aus den physischen Eigenschaften ergeben, nicht aber anders herum auch die physischen Eigenschaften aus den nicht-physischen. Die nicht-physischen Eigenschaften sind kraft der physischen Eigenschaften instantiiert, aber die physischen Eigenschaften sind nicht kraft der nicht-physischen Eigenschaften instantiiert. Die physischen Eigenschaften haben ontologische Priorität. Eine solche Priorität kann aber nicht in beiden Richtungen bestehen.

Zweitens sollten sich die nicht-physischen Eigenschaften mit einer gewissen modalen Stärke aus den physischen Eigenschaften ergeben. Dabei reicht es plausiblerweise nicht, dass bloß naturgesetzliche Zusammenhänge zwischen physischen und nicht-physischen Eigenschaften bestehen. Gott konnte sich nach der Erschaffung der physischen Eigenschaftsverteilung nicht deshalb zurücklehnen, weil er zuvor explizite Naturgesetze erlassen hat, die nicht-physische Eigenschaften mit physischen Eigenschaften verbinden. Er konnte sich zurücklehnen, weil solche Naturgesetze gar nicht nötig sind. Gegeben die physischen Eigenschaften (und die physischen Naturgesetze) konnte es, so die Idee, *nicht einmal Gott verhindern*, dass die nicht-physischen Eigenschaften sich ergeben. Die nicht-physischen Eigenschaften ergeben sich mit *metaphysischer Notwendigkeit* (und

[88] Vgl. für Formulierungen des Physikalismus, die sich direkt auf solcherlei Behauptungen berufen, z. B. Nimtz (2009) und Wilson (2016).

nicht bloß mit nomologischer Notwendigkeit) aus den physischen Eigenschaften und Gesetzen.[89]

Hieraus ergibt sich drittens, dass es sich bei der gesuchten Beziehung nicht um eine kausale Beziehung handelt. Kausale Beziehungen beruhen in der einen oder anderen Weise auf naturgesetzlichen Zusammenhängen. Sie verknüpfen zudem gänzlich distinkte Ereignisse. Nicht-reduktive Physikalist*innen sind hingegen auf der Suche nach einer Relation zwischen Eigenschaften, die mit metaphysischer Notwendigkeit einhergeht und die es plausibel macht, dass nicht-physische Eigenschaften nicht mehr als physische Eigenschaften (und in diesem Sinne nicht gänzlich distinkt von physischen Eigenschaften) sind.

Kurzum: Nicht-reduktive Physikalist*innen sind auf der Suche nach einer asymmetrischen (oder zumindest nicht-symmetrischen), nicht-kausalen und metaphysisch notwendigen Abhängigkeitsbeziehung zwischen nicht-physischen und physischen Eigenschaften, die plausibilisiert, dass nicht-physische Eigenschaften nicht mehr als physische Eigenschaften sind und aufgrund von physischen Eigenschaften instantiiert sind.

2.5.2 Supervenienz

Eine Art, der theologischen Metapher einen greifbareren Inhalt zu geben, beruht auf Supervenienz-Beziehungen. Der Supervenienz-Ansatz der hinreichend starken ontologischen Abhängigkeit besagt Folgendes:

---

[89] Dies setzt natürlich voraus, dass es eine geeignete Unterscheidung zwischen metaphysischer und nomologischer Modalität gibt. Ich gehe im Folgenden von der verbreiteten Idee aus, dass die Menge der nomologisch möglichen Welten durch die aktualen Naturgesetze festgelegt ist: Nomologisch mögliche Welten sind Welten, in denen dieselben Naturgesetze gelten wie in der aktualen Welt. Zudem gehe ich davon aus, dass die Menge nomologisch möglicher Welten eine echte Teilmenge der Menge der metaphysisch möglichen Welten ist. Es gibt demnach metaphysisch mögliche Welten, in denen die aktualen Naturgesetze nicht gelten und die entsprechend nomologisch unmöglich sind. In diesem Sinne ist metaphysische Notwendigkeit stärker als nomologische Notwendigkeit: Metaphysische Notwendigkeit impliziert nomologische Notwendigkeit, aber nomologische Notwendigkeit impliziert nicht metaphysische Notwendigkeit. Auf die Konsequenzen der alternativen Auffassung, dass die aktualen Naturgesetze metaphysisch notwendig sind, für die Formulierung des nicht-reduktiven Physikalismus gehe ich in Abschnitt 2.6.6. noch einmal ein. Vgl. auch Kment (2017) für einen Überblick über die Problematik des Verhältnisses zwischen verschiedenen Arten von Modalität..

> **(OntA-Sup)** Nicht-physische Eigenschaften sind genau dann in einem hinreichend starken Sinne von physischen Eigenschaften ontologisch abhängig, wenn gilt: Nicht-physische Eigenschaften supervenieren auf physischen Eigenschaften.

Die Grundidee hinter Supervenienz-Beziehungen kann durch die folgende vorläufige Bestimmung auf den Punkt gebracht werden:

> **(Sup-Gru)** Nicht-physische Eigenschaften supervenieren genau dann auf physischen Eigenschaften, wenn gilt: Objekte, die sich in ihren physischen Eigenschaften gleichen, müssen sich auch in ihren nicht-physischen Eigenschaften gleichen.

Nun lassen sich von dieser Bestimmung ausgehend eine Reihe von unterschiedlichen Supervenienz-Beziehungen definieren, die sich durch die betroffenen Objekte und durch ihre modale Stärke unterscheiden.

Zunächst einmal lässt sich auch eine globale Supervenienz-These aufstellen, die nicht einzelne Objekte, sondern ganze Welten betrifft. Die Idee ist hier also, dass sich Welten, welche sich in der in ihnen vorkommenden Verteilung physischer Eigenschaften gleichen, auch in ihrer Verteilung der restlichen Eigenschaften gleichen. Hieraus ergibt sich eine recht direkte Ausdeutung der anfangs genannten Metapher: Gott muss die nicht-physischen Eigenschaften nicht extra erschaffen, weil sie auf den physischen Eigenschaften global supervenieren: Eine Welt, die diese Verteilung von physischen Eigenschaften hat, muss auch jene Verteilung von nicht-physischen Eigenschaften haben.

Eine solche globale Supervenienz-These wird häufig verwendet, um den Physikalismus zu definieren. So bestimmt Frank Jackson den globalen Physikalismus wie folgt:

> **(GP-J)** Der globale Physikalismus ist genau dann wahr, wenn gilt: Jedes minimale physische Duplikat der aktualen Welt ist ein Duplikat der aktualen Welt *simpliciter*.[90]

Hiermit ist nicht die These verbunden, dass die aktuale Welt keine nicht-physischen Eigenschaften enthält. Es ist damit aber die These verbunden, dass

[90] Ein minimales physisches Duplikat der aktualen Welt ist eine Welt, die genau dieselbe physische Eigenschaftsverteilung wie die aktuale Welt hat, und die darüber hinaus keine anderen Eigenschaften hat, die unsere Welt nicht hat. Vgl. Jackson (1998: 12).

die nicht-physischen Eigenschaften, die in der aktualen Welt instantiiert sind, auch in jeder Welt instantiiert sind, die ein minimales physisches Duplikat der aktualen Welt ist. Die These, dass alle weiteren Eigenschaften in unserer Welt auf den physischen Eigenschaften supervenieren, läuft dann darauf hinaus, dass keine Welt, die dieselbe Verteilung physischer Eigenschaften hat, sich in der Verteilung irgendwelcher anderer Eigenschaften unterscheiden kann: Jede mögliche Welt, die der aktualen Welt in physischer Hinsicht bis ins Letzte gleicht, gleicht der aktualen Welt auch in jeder anderen Hinsicht bis ins Letzte.

Die Bestimmung des Physikalismus über die These der globalen Supervenienz verweist auf eine wichtige Gemeinsamkeit zwischen reduktiven und nicht-reduktiven Versionen des Physikalismus: Sowohl reduktive als auch nicht-reduktive Physikalist*innen akzeptieren die These der globalen Supervenienz aller Eigenschaften auf den physischen Eigenschaften. Die These der globalen Supervenienz drückt somit eine grundlegende und allgemein akzeptierte Verpflichtung jeder Form des Physikalismus aus. Im reduktiven Physikalismus ergibt sich die globale Supervenienz jedoch trivialerweise aus dem Umstand, dass es keine nicht-physischen Eigenschaften gibt. Im nicht-reduktiven Physikalismus hingegen drückt die These der globalen Supervenienz eine wichtige Bedingung an die Zulässigkeit nicht-physischer Eigenschaften aus: Nur solche nicht-physischen Eigenschaften sind in der Ontologie des nicht-reduktiven Physikalismus zulässig, die auf den physischen Eigenschaften global supervenieren.

Dies ist eine wichtige Bedingung, aber nicht die einzige wichtige Bedingung. Die These der globalen Supervenienz betrifft die komplette Verteilung von physischen und nicht-physischen Eigenschaften in unserer Welt. Sie ist aber blind für lokale Ausbildungen der globalen Eigenschaftsverteilung. Beispielsweise ist es mit der These der globalen Supervenienz vereinbar, dass sich in unserer Welt zwei Personen in ihren physischen Eigenschaften exakt gleichen, sie sich in ihren mentalen Eigenschaften aber radikal unterscheiden: So könnte eine Person glauben, dass Paris in Frankreich ist, während ihr physischer Doppelgänger in derselben Welt gar nichts glaubt, oder glaubt, dass Paris in Deutschland ist usw. Entsprechendes gilt auch schon für z. B. biologische Eigenschaften. So könnten sich zwei Gene in unserer Welt in ihren physischen Eigenschaften exakt gleichen, aber gänzlich unterschiedliche Erbinformationen tragen. Solange diese mentalen bzw. biologischen Unterschiede sich auch in minimalen physischen Duplikaten der Welt wiederfinden, ist der These der globalen Supervenienz Genüge getan.[91]

[91] Vgl. Petrie (1987: 120–122) und Kim (1987: 319–321). Dies zeigt auch, dass globale Supervenienz – zumindest in der hier zitierten Version von Jackson – nicht mit starker oder schwacher lokaler Supervenienz äquivalent ist.

Es ist daher sinnvoll, die Zulässigkeit nicht-physischer Eigenschaften im nicht-reduktiven Physikalismus weiter einzuschränken. Auch hier können Supervenienz-Thesen hilfreich sein. Diese sollten sich jedoch nicht auf ganze Welten beziehen, sondern auf einzelne Objekte. Ein Beispiel für eine solche lokale, objektbezogene Supervenienz-Beziehung ist die folgende:

> **(Sup-L-Schwach)** Nicht-physische Eigenschaften supervenieren genau dann lokal schwach auf physischen Eigenschaften, wenn gilt: Für jede mögliche Welt w und jedes Objekt x und y in w gilt: wenn x und y dieselben physischen Eigenschaften haben, dann haben sie dieselben nicht-physischen Eigenschaften.[92]

Diese lokale Supervenienz-Beziehung verbietet, dass es in unserer Welt zwei Objekte gibt, die sich in ihren physischen Eigenschaften gleichen, aber in ihren nicht-physischen Eigenschaften unterscheiden. Die oben genannten Fälle sind daher ausgeschlossen, wenn wir das Bestehen einer solchen lokalen Supervenienz-Beziehung als weitere Bedingung für zulässige nicht-physische Eigenschaften einführen. Die These verbietet aber noch mehr: Es ist *auch in anderen möglichen Welten* ausgeschlossen, dass sich zwei Objekte, die sich ‚dort' in ihren physischen Eigenschaften gleichen, in ihren nicht-physischen Eigenschaften unterscheiden. Das heißt: dies kommt nicht nur (zufälligerweise) in unserer Welt nicht vor, sondern es kommt auch in anderen möglichen Welten nicht vor. Für jede mögliche Welt gilt: Zwei Objekte in dieser Welt gleichen sich in ihren

---

Es gibt eine Strategie, entgegen dieser Überlegung zu argumentieren, dass globale Supervenienz doch starke lokale Supervenienz impliziert: Man kann sich nämlich auf ein Isolationsprinzip mit Blick auf mögliche Welten beziehen: Insofern dann eine Welt möglich ist, die von Peter und seinem physischen Zwilling bevölkert ist und in der sich Peter und sein physischer Zwilling in ihren mentalen Eigenschaften unterscheiden, ist auch eine Welt möglich, die nur von Peter bevölkert ist und in der Peter allein mit seiner mentalen Eigenschaft auftaucht und eine andere Welt möglich, die nur von Peters Zwilling bevölkert ist und in der Peters Zwilling allein mit seiner anderen mentalen Eigenschaft auftaucht. Diese Kombination von möglichen Welten verletzt dann globale Supervenienz. Unter dieser Voraussetzung folgt aus der hier geschilderten Verletzung starker und schwacher lokaler Supervenienz also eine Verletzung globaler Supervenienz (vgl. Paull und Sider (1992)).

Stalnaker (1996) zeigt zudem, dass globale und starke Supervenienz äquivalent sind, wenn die Klasse der Eigenschaften in der Supervenienzbasis unter verschiedenen eigenschaftsformenden Operationen (wie Negation, Konjunktion und Disjunktion) geschlossen ist. Vgl. auch Hoffmann und Newen (2007) für weitere Diskussion.

Ich danke Vera Hoffmann-Kolss für hilfreiche Hinweise zu dieser Diskussion.

[92] Vgl. Abschnitt 4.1. in McLaughlin und Bennett (2018) und Kim (1984a: 58).

nicht-physischen Eigenschaften, wenn sie sich in ihren physischen Eigenschaften gleichen.

Allerdings werden durch diese Supervenienz-Beziehung nur Objekte verglichen, die sich innerhalb derselben Welt befinden. Es ist daher vereinbar mit dem Bestehen der Beziehung, dass es in sehr nahen möglichen Welten Objekte gibt, die dieselben physischen Eigenschaften haben wie ein Objekt in der aktualen Welt, die aber gänzlich andere nicht-physische Eigenschaften haben. Es ist mit dem Bestehen der Beziehung zum Beispiel vereinbar, dass es sehr leicht hätte sein können, dass ein Gen dieselben physischen Eigenschaften hat, die es tatsächlich hat, aber gänzlich andere Erbinformationen trägt. Dementsprechend lassen sich durch diesen Zusammenhang auch keine kontrafaktischen Konditionale der Art ‚Wenn a andere nicht-physische Eigenschaften gehabt hätte, dann hätte a auch andere physische Eigenschaften gehabt' begründen. Die Idee, dass es keine nicht-physischen Unterschiede ohne physische Unterschiede geben *kann* und dass sich zwei Objekte, die sich in ihren physischen Eigenschaften gleichen, auch in ihren nicht-physischen Eigenschaften gleichen *müssen*, ist also nicht zufriedenstellend eingefangen.

Diese Beobachtung motiviert die Formulierung einer stärkeren lokalen Supervenienz-Beziehung:

> **(Sup-L-Stark)** Nicht-physische Eigenschaften supervenieren genau dann lokal stark auf physischen Eigenschaften, wenn gilt: Für alle möglichen Welten *w* und *w*' und jedes Objekt *o* in *w* und *o*' in *w*' gilt: Wenn *o* in *w* dieselben physischen Eigenschaften hat wie *o*' in *w*', dann haben *o* und *o*' dieselben nicht-physischen Eigenschaften.[93]

Hier werden nun auch Objekte verglichen, die sich in unterschiedlichen möglichen Welten befinden: Es kann gar nicht sein, dass ein Objekt, das dieselben physischen Eigenschaften hat wie ein bestimmtes Objekt in unserer Welt, andere nicht-physische Eigenschaften hat. Entsprechend folgt auch ein kontrafaktisches Konditional der Art ‚Wenn a andere nicht-physische Eigenschaften gehabt hätte, dann hätte a auch andere physische Eigenschaften gehabt'. Der enge modale Zusammenhang zwischen physischen Eigenschaften und nicht-physischen Eigenschaften wird also besser eingefangen. Der nicht-reduktive Physikalismus ist nicht nur auf die These der globalen Supervenienz festgelegt, sondern auch auf die stärkere These, dass nicht-physische Eigenschaften lokal stark auf physischen Eigenschaften supervenieren.

---

[93] Vgl. Abschnitt 4.1. in McLaughlin und Bennett (2018) sowie Kim (1984a: 65).

Der Unterschied zwischen schwacher und starker Supervenienz betrifft die Frage, ob bloß Objekte innerhalb einer Welt verglichen werden oder auch Objekte, die sich in unterschiedlichen Welten befinden. Starke Supervenienz ist in dem Sinne modal stärker, dass auch weltübergreifende Vergleiche relevant sind. Eine andere Frage zur modalen Stärke wurde jedoch bis hierher nicht berührt: Über welche möglichen Welten wird in den obigen Definitionen quantifiziert? Über alle metaphysisch möglichen Welten? Oder über alle nomologisch möglichen Welten?

Ich gehe davon aus, dass es im Kontext des nicht-reduktiven Physikalismus auf den metaphysischen Sinn der Modalität ankommt. Das heißt, dass die obigen Definitionen so zu verstehen sind, dass über alle metaphysisch möglichen Welten quantifiziert wird. Wenn es bloß um naturgesetzliche Zusammenhänge ginge, wäre eine Verbindung zu der Idee, dass nicht-physische Eigenschaften *nicht mehr als* physische Eigenschaften sind, von vornherein nicht plausibel.[94]

Nicht-reduktive Physikalist*innen sind also darauf festgelegt, dass nicht-physische Eigenschaften mit metaphysischer Notwendigkeit lokal stark auf physischen Eigenschaften supervenieren.[95] Hieraus folgt, dass sie auch auf die entsprechende schwache lokale und auf die globale Supervenienz-These festgelegt sind, da diese Supervenienz-Thesen aus der These der starken lokalen Supervenienz folgen.

Es ist jedoch fraglich, ob sich nicht-reduktive Physikalist*innen mit einer solchen Charakterisierung der Beziehung zwischen physischen und nicht-physischen Eigenschaften schon zufriedengeben kann. Sämtliche Supervenienz-Beziehungen drücken zunächst nur ein bestimmtes modales Muster der Kovariation von Eigenschaftsmengen aus. Sie erklären aber nicht, wie es zu dieser Kovariation kommt und warum diese Kovariation besteht.[96] Dies sieht man schon daran, dass die Supervenienz-Thesen sämtlich auch von reduktiven Physikalist*innen akzeptiert werden. Tatsächlich ergibt sich ihre Wahrheit im reduktiven Physikalismus trivialerweise daraus, dass es keine nicht-physischen Eigenschaften gibt bzw. dass

[94] Vgl. z. B. Bennett (2008: 284–286).

[95] Kim (2005: 49) meint (überraschend), dass "we can set aside the possibility that mind-body supervenience is logically or metaphysically necessary, since such a view is essentially a reductionist view". Auch Bennett (2008: 286) wundert sich über diesen Kommentar. Klar ist zumindest, dass metaphysisch notwendige Supervenienz mit der Ablehnung von Eigenschaftsidentitäten und in diesem Sinne mit einem Nicht-Reduktionismus kompatibel ist. Besonders rätselhaft ist der Kommentar, wenn man ihn einem Kommentar aus Kim (2002a: 9) gegenüberstellt: "I believe that emergentists would have accepted mind-body supervenience in a very strong sense, with necessity interpreted as metaphysical necessity".

[96] Vgl. Horgan (1993).

spezialwissenschaftliche und mentale Eigenschaften mit physischen Eigenschaften identisch sind. Reduktive Physikalist*innen haben also einen guten Grund, die Supervenienz-Thesen anzuerkennen. Dies liegt daran, dass sie eine überzeugende Erklärung für das Bestehen der Supervenienz-Beziehung haben.

Da nicht-reduktive Physikalist*innen die Existenz nicht-physischer, mentaler und spezialwissenschaftlicher Eigenschaften anerkennen und Eigenschaftsidentitäten ablehnen, steht ihnen *diese* Erklärung der Supervenienz-Thesen nicht zur Verfügung. Sie stehen daher vor der Herausforderung, eine alternative Erklärung anzubieten. Die bloße Ablehnung von Eigenschaftsidentitäten leistet hier keine Erklärungsarbeit. Es gilt daher, die Beziehung zwischen physischen Eigenschaften auf eine Weise positiv zu bestimmen, die *erklärt*, warum nicht-physische Eigenschaften auf physischen Eigenschaften supervenieren.

Eng hiermit zusammen hängt der folgende Punkt: Supervenienz ist selbst keine *metaphysisch erklärende* Relation. Man kann daraus, dass eine Eigenschaftsklasse auf einer anderen Eigenschaftsklasse superveniert, nicht ableiten, dass die supervenierenden Eigenschaften instantiiert sind, *weil* die Eigenschaften in der Supervenienzbasis instantiiert sind, oder dass die supervenierenden Eigenschaften *kraft* der Eigenschaften in der Supervenienzbasis instantiiert sind. Entsprechend etabliert die bloße Supervenienz mentaler Eigenschaften auf physischen Eigenschaften noch nicht, dass mentale Eigenschaften *aufgrund von* physischen Eigenschaften instantiiert sind.

Das sieht man unter anderem daran, dass notwendig instantiierte Eigenschaften trivialerweise auf beliebigen anderen Eigenschaften supervenieren. Die Eigenschaft, rot-oder-nicht-rot zu sein, superveniert zum Beispiel auf Form-Eigenschaften (oder auf jeder beliebigen anderen Klasse von Eigenschaften). Zwei Objekte, die sich in ihren Form-Eigenschaften gleichen (z. B. beide rund usw. sind), müssen sich ganz einfach deshalb auch darin gleichen, rot-oder-nicht-rot zu sein, weil sich alle Objekte notwendigerweise darin gleichen, rot-oder-nicht-rot zu sein. Hieraus lässt sich aber sicherlich nicht folgern, dass irgendwelche Objekte rot-oder-nicht-rot sind, *weil* sie z. B. rund sind.[97]

Insofern durch die Relation der hinreichend starken ontologischen Abhängigkeit aber auch solche Behauptungen der metaphysischen Erklärung ergeben sollten, kann die Supervenienz-Relation die Rolle der ontologischen Abhängigkeit im nicht-reduktiven Physikalismus also nicht lückenlos übernehmen. Entsprechend wäre es zumindest wünschenswert, wenn nicht-reduktive Physikalist*innen sich auf eine Relation ontologischer Abhängigkeit berufen könnten, die einerseits

[97] Vgl. auch Kapitel 3in McLaughlin und Bennett (2018).

erklärt, warum mentale und spezialwissenschaftliche Eigenschaften auf physischen Eigenschaften supervenieren und andererseits etabliert, dass mentale und spezialwissenschaftliche Eigenschaften *aufgrund von* physischen Eigenschaften instantiiert sind.

2.5.3 Funktionale Realisierung

Ein Ansatz, dies zu leisten, beruht auf einer funktionalistischen Theorie mentaler und spezialwissenschaftlicher Eigenschaften. Die Grundidee dieser Ansätze besteht darin, dass die entsprechenden Eigenschaften durch ihre kausale Rolle bestimmt sind: Die Eigenschaften (oder genauer die Ereignisse, für die diese Eigenschaften konstitutiv sind) haben charakteristische Ursachen und Wirkungen. Dies macht ihre kausale Rolle aus. Als Standardbeispiel kann die verbreitete, vereinfachte, aber veranschaulichende Bestimmung der Eigenschaft, Schmerz zu haben, über dessen kausale Rolle dienen: Schmerz wird typischerweise durch Verletzungen verursacht und verursacht typischerweise die Überzeugung, dass man eine Verletzung hat und bestimmtes Verhalten, wie Stöhnen oder Aufschreien: Er wird durch bestimmte physische oder mentale kausale Inputs verursacht und bewirkt bestimmte physische oder mentale Outputs. Dies macht die kausale Rolle von Schmerz aus.[98]

Die Idee ist nun, dass Schmerz über diese kausale Rolle bestimmt werden kann. Eine Art und Weise, diese Idee zu präzisieren, besteht darin, Schmerz als eine Eigenschaft zweiter Ordnung zu bestimmen: Schmerz ist die Eigenschaft, eine Eigenschaft zu haben, die die Schmerz-Rolle hat. Schmerz ist in diesem Bild nicht selbst der Inhaber der Schmerz-Rolle. Stattdessen gibt es distinkte Eigenschaften, zum Beispiel bestimmte neurologische Eigenschaften, die nicht identisch mit Schmerz sind und die (in unterschiedlichen Lebewesen oder unterschiedlichen Situationen) die Schmerz-Rolle haben. Wer eine solche Eigenschaft hat, hat Schmerz. Man sagt dann, dass die Eigenschaft, die die Schmerz-Rolle hat, Schmerz *realisiert*.

Realisierung ist demnach eine Beziehung, die zwischen Eigenschaften erster Ordnung und Eigenschaften zweiter Ordnung besteht. Die Eigenschaft zweiter Ordnung ist im Allgemeinen die Eigenschaft, eine Eigenschaft mit der kausalen Rolle R zu haben. Eine Eigenschaft erster Ordnung realisiert diese Eigenschaft

[98] Vgl. Block (1980), Levin (2018). Es gibt Alternativen zu dieser Auffassung von Realisierung, auf die ich hier nicht detailliert eingehe. Vgl. insbesondere Shoemaker (2007) sowie Gillett (2003).

zweiter Ordnung genau dann, wenn sie die kausale Rolle R hat. Die Eigenschaft zweiter Ordnung ist genau dann instantiiert, wenn sie durch irgendeine Eigenschaft erster Ordnung realisiert ist.[99]

Diese Realisierungsrelation erlaubt die Möglichkeit multipler Realisierung: Es kann viele unterschiedliche physische Eigenschaften geben, die eine bestimmte kausale Rolle spielen und die sich (von ihrer kausalen Rolle abgesehen) in keiner Weise ähneln. Es ist zum Beispiel ohne weiteres möglich, dass sowohl Oktopusse als auch Menschen in dem Sinne Schmerzen haben, dass sie eine Eigenschaft haben, die die Schmerz-Rolle hat –, dass aber neuronale Eigenschaft $\mathbf{N_1}$ in Menschen die Schmerz-Rolle hat, während neuronale Eigenschaft $\mathbf{N_2}$ in Oktopussen die Schmerzrolle hat. Hierfür muss nicht vorausgesetzt werden, dass $\mathbf{N_1}$ und $\mathbf{N_2}$ neben ihres Habens der Schmerz-Rolle durch weitere Gemeinsamkeiten verbunden sind. Oktopusse und Menschen müssen also nicht eine neuronale (oder gar physische$_{eng}$) Eigenschaft teilen, um eine mentale Eigenschaft teilen zu können.

Zugleich ergibt sich aus der Relation der funktionalen Realisierung ein klarer Sinn, in dem mentale und spezialwissenschaftliche Eigenschaften *aufgrund von* physischen Eigenschaften instantiiert sind: Susi hat deshalb die Eigenschaft, eine Eigenschaft zu haben, die die Schmerz-Rolle hat, *weil* sie eine Eigenschaft hat, die die Schmerz-Rolle hat. Die Eigenschaft zweiter Ordnung ergibt sich aus der Eigenschaft erster Ordnung. Insofern die Realisierer mentaler Eigenschaften sämtlich physisch sind, sind mentale Eigenschaften also aufgrund von physischen Eigenschaften instantiiert.

In welchem Verhältnis steht Realisierung zu Supervenienz? Einige Autor*innen meinen, dass die These, dass mentale und spezialwissenschaftliche Eigenschaften funktionale Eigenschaften sind, die durch physische Eigenschaften realisiert sind, *erklärt*, warum mentale und spezialwissenschaftliche Eigenschaften auf physischen Eigenschaften supervenieren.[100] Gegenstände, die sich in ihren physischen Eigenschaften gleichen, können sich in ihren mentalen und spezialwissenschaftlichen Eigenschaften nicht unterscheiden, weil die physischen Eigenschaften der Gegenstände auch die physischen Realisierer mentaler und spezialwissenschaftlicher Eigenschaften umfassen. Und wenn ein Gegenstand einen

[99] Dieser 'Rollen-Funktionalismus', in dem mentale Zustände mit funktionalen Eigenschaften zweiter Ordnung identifiziert werden, wird üblicherweise mit einem 'Füller-Funktionalismus' kontrastiert, in dem mentale Zustände mit den Eigenschaften erster Ordnung identifiziert werden, die die entsprechenden Rollen haben. Der Füller-Funktionalismus ist eine reduktive Theorie und stößt auf die üblichen Probleme im Zusammenhang mit multipler Realisierung. Vgl. hierzu z. B. Abschnitt 3.4. in Levin (2018).

[100] Vgl. z. B. Melnyk (2003a: 52).

Realisierer einer mentalen Eigenschaft instantiiert, dann muss er auch die zugehörige mentale Eigenschaft instantiieren. Insofern zwei Gegenstände also dieselben physischen Realisierer mentaler Eigenschaften haben, müssen sie auch dieselben mentalen Eigenschaften haben.

Diese Argumentation beruht jedoch auf der Annahme, dass die Realisierer mentaler Eigenschaften ihre zugehörigen mentalen Eigenschaften necessitieren. Dies ist jedoch nicht sehr plausibel: Denn die physischen Eigenschaften, die *tatsächlich* Schmerz realisieren, tun dies nicht notwendigerweise. In anderen möglichen Welten (oder auch schon in anderen Situationen in der aktualen Welt) haben sie die kausale Rolle, die sie tatsächlich haben, nicht. Es ist möglich, dass die Eigenschaft $\mathbf{N_1}$, die in der aktualen Welt die für Schmerz typischen Ursachen und Wirkungen hat, in anderen möglichen Welten andere Ursachen und Wirkungen hat. In diesen Welten kann der Realisierer $\mathbf{N_1}$ instantiiert sein, ohne dass Schmerz instantiiert ist. Also kann man nicht sagen, dass $\mathbf{N_1}$ Schmerz necessitiert.

Besonders einsichtig wird dies, wenn man sich an den typischen – halb fiktiven oder zumindest stark idealisierten – Beispielen für Realisierer orientiert. Als Realisierer von Schmerz wird beispielhaft häufig die neurologische Eigenschaft, ein C-Faser Feuern zu sein, ins Spiel gebracht. C-Faser Feuern könnte jedoch auch an isolierten C-Fasern in einem Reagenzglas instantiiert sein. Wenn C-Faser Feuern unter solchen Umständen instantiiert ist, wird man nicht erwarten, dass C-Faser-Feuern (in einer solchen Situation) Schmerz realisiert.[101] Denn das C-Faser Feuern wird dann weder von Verletzungen verursacht noch verursacht es Aufschreien o.ä. Es hat dann eben nicht die für Schmerz charakteristische kausale Rolle.

Diese Beobachtung motiviert eine Unterscheidung zwischen Kern-Realisierern und totalen Realisierern.[102] Kern-Realisierer sind dabei typischerweise relativ ‚lokale' Eigenschaften wie C-Faser-Feuern. Sie spielen (aktual) die kausale Rolle, durch die die zugehörige mentale Eigenschaft charakterisiert ist. Allerdings necessitieren sie die zugehörige mentale Eigenschaft für sich genommen nicht. Es ist also möglich, dass ein Kern-Realisierer ohne die zugehörige mentale Eigenschaft instantiiert ist.

Vielmehr müssen gewisse Umstände – weitere physische Eigenschaftsinstantiierungen und das Gelten physischer Gesetze – hinzukommen, um zu einer – dann recht komplexen – physischen Eigenschaft zu gelangen, die eine mentale Eigenschaft metaphysisch necessitiert. Diese komplexe physische Eigenschaft – die aus dem Kern-Realisierer plus den zusätzlich nötigen Umständen besteht – kann

[101] Vgl. Bennett (2003: 485) und Shoemaker (2007: 21).

[102] Die Unterscheidung wird von Shoemaker (1981: 96) eingeführt.

dann als totaler Realisierer bezeichnet werden. Der totale Realisierer einer mentalen Eigenschaft necessitiert die mentale Eigenschaft: Wenn der totale Realisierer instantiiert ist, muss die zugehörige mentale Eigenschaft ebenfalls instantiiert sein.

Eine Erklärung der Supervenienz mentaler Eigenschaften auf physischen Eigenschaften ist nur durch die Relation der totalen Realisierung möglich. Wenn jede mentale Eigenschaft einen physischen *totalen* Realisierer hat, ist auch klar, warum mentale Eigenschaften auf physischen Eigenschaften supervenieren. Zwei Objekte mit denselben physischen Eigenschaften haben dann dieselben totalen Realisierer mentaler Eigenschaften. Deshalb haben sie dann auch dieselben mentalen Eigenschaften. Und deshalb können sie sich nicht in ihren mentalen Eigenschaften unterscheiden, ohne sich auch in ihren physischen Eigenschaften zu unterscheiden. Zugleich ist mit totaler Realisierung auch zumindest ein Ansatz für die Erklärung der modalen Verknüpfung zwischen totalen Realisierern und den zugehörigen mentalen Eigenschaften gegeben: Totale Realisierer necessitieren ihre zugehörigen mentalen Eigenschaften, weil sie es notwendig machen, dass eine physische Eigenschaft instantiiert ist, die die für die mentale Eigenschaft charakteristische kausale Rolle spielt.

2.5.4 Determination

Ein weiterer Ansatz, zu erklären, warum mentale und spezialwissenschaftliche Eigenschaften auf physischen Eigenschaften supervenieren und deutlich zu machen, inwiefern mentale und spezialwissenschaftliche Eigenschaften aufgrund von physischen Eigenschaften instantiiert sind, besteht in der Behauptung, dass mentale Eigenschaften Determinablen sind, deren Determinaten physische Eigenschaften sind.[103] Unter ‚Determination' wird hier also die Beziehung zwischen Determinaten und Determinablen verstanden.

Ein häufig bemühtes Beispiel für so verstandene Determination ist das folgende: Die Eigenschaft, rot zu sein, ist eine Determinable. Es gibt verschiedene spezifischere Arten und Weisen, wie ein Objekt rot sein kann. Zum Beispiel können Objekte scharlachrot oder bordeauxrot sein. Die Eigenschaft, scharlachrot zu sein und die Eigenschaft, bordeauxrot zu sein, sind zwei Determinaten der Determinablen, rot zu sein. Beide Eigenschaften *determinieren* also jeweils die Eigenschaft, rot zu sein.

[103] Vgl. z. B. Macdonald und Macdonald (1986), Yablo (1992) und Wilson (2009).

Determinaten sind im Allgemeinen *spezifischer* als ihre Determinablen. Determinaten sind verschiedene spezifischere Arten und Weisen, wie etwas die zugehörige Determinable haben kann. Scharlachrot zu sein und bordeauxrot zu sein sind zwei Arten, wie etwas rot sein kann. Dieselbe Eigenschaften kann dabei sowohl Determinable mit Blick auf spezifischere Eigenschaften als auch Determinate mit Blick auf weniger spezifische Eigenschaften sein. Die Eigenschaft, rot zu sein, ist mit Blick auf die Eigenschaft, scharlachrot zu sein und die Eigenschaft, bordeauxrot zu sein, eine Determinable. Mit Blick auf die Eigenschaft, farbig zu sein, ist die Eigenschaft, rot zu sein, hingegen eine Determinate – neben anderen Determinaten wie der Eigenschaft blau zu sein oder der Eigenschaft, grün zu sein.

Die so verstandene Relation der Determination hat einige Merkmale, die sie auf den ersten Blick gut dazu geeignet erscheinen lassen, die Beziehung zwischen physischen und mentalen Eigenschaften im nicht-reduktiven Physikalismus zu erhellen:

Determination ist erstens – allerdings mit gewissen Einschränkungen, auf die ich gleich noch eingehe – mit multipler Realisierung[104] vereinbar: Verschiedene Objekte können dieselbe Determinable instantiieren, ohne aber eine bestimmte Determinate zu teilen. Dieser Ball und jenes T-Shirt sind beide rot. Allerdings ist der Ball scharlachrot und das T-Shirt bordeauxrot. Also ist es nicht der Fall, dass Determinablen eine bestimmte Determinate necessitieren. Ebenso ist es laut der These der multiplen Realisierung nicht der Fall, dass mentale Eigenschaften eine bestimmte physische Eigenschaft necessitieren. Die Annahme, dass mentale Eigenschaften Determinablen sind, deren Determinaten physische Eigenschaften sind, scheint also zumindest vereinbar mit der multiplen Realisierung mentaler Eigenschaften zu sein.

Zweitens ist es anders herum der Fall, dass Determinaten ihre Determinable metaphysisch necessitieren: Wenn ein Objekt scharlachrot ist, dann muss es auch rot sein. Die Annahme, dass mentale Eigenschaften Determinablen mit physischen Determinaten sind, liefert also außerdem eine gute Erklärung dafür, dass physische Eigenschaften die zugehörigen mentalen Eigenschaften necessitieren. Hierdurch wird zugleich eine Erklärung dafür gegeben, dass mentale Eigenschaften auf physischen Eigenschaften supervenieren: Zwei Objekte, die sich in ihren physischen Eigenschaften gleichen, können sich in ihren mentalen Eigenschaften

[104] ‘Realisierung’ in ‘multiple Realisierung’ wird hier wieder in einem unspezifischen Sinne verstanden. Es ist hiermit nicht die Behauptung verbunden, dass mentale Eigenschaften durch physische Eigenschaften *im Sinne der in 2.5.3. spezifizierten Realisierungsrelation* realisiert sind.

deshalb nicht unterscheiden, weil sie dieselben physischen Determinaten mentaler Eigenschaften haben. Schließlich können sich auch zwei scharlachrote Objekte nicht in ihrer Eigenschaft, rot zu sein, unterscheiden.

Drittens ist Determination plausiblerweise eine Relation, die ersichtlich macht, inwiefern Determinablen *aufgrund von* ihren Determinaten instantiiert sind. Dieser Ball ist rot, *weil* er scharlachrot ist. Entsprechend würde die Annahme, dass mentale Eigenschaften Determinablen mit physischen Determinaten sind, auch einen Schluss darauf erlauben, dass mentale Eigenschaften *aufgrund von* physischen Eigenschaften instantiiert sind.

Gegeben diese Merkmale der Determinationsrelation scheint die Annahme, dass mentale Eigenschaften Determinablen mit physischen Determinaten sind, also eine Reihe von Desiderata ontologischer Abhängigkeit im Kontext einer Formulierung des nicht-reduktiven Physikalismus auf einfache und einheitliche Weise zu erfüllen.

Auf der anderen Seite hat die Relation der Determination jedoch auch einige Merkmale, die es schwierig erscheinen lassen, die Beziehung zwischen physischen und mentalen Eigenschaften als eine Determinationsbeziehung zu verstehen:

Die Determinaten einer Determinablen sind entlang von Unterscheidungsdimensionen voneinander unterschieden, die einen Bezug zur Determinablen haben.[105] Verschiedene Rottöne unterscheiden sich mit Blick auf ihre Helligkeit oder Sättigung in ihrer Art, rot zu sein. Auch bei mentalen Eigenschaften wie Schmerz lassen sich solche Unterscheidungsdimensionen angeben, anhand derer spezifischere *mentale* Determinaten einer mentalen Determinablen unterschieden werden können. So können sich Schmerzen etwa in ihrer Intensität, ihrer Lokalisation oder ihrer phänomenalen Qualität unterscheiden: Intuitiv plausible mentale Determinaten der Determinablen Schmerz sind also die spezifischeren mentalen Eigenschaften, einen starken, stechenden Kopfschmerz oder einen leichten, drückenden Rückenschmerz zu haben.[106]

Nun ist der hier diskutierte Vorschlag jedoch, dass mentale Eigenschaften *physische* Determinaten haben. Die komplexen physischen Eigenschaften, die Schmerz necessitieren, sollen Determinaten von Schmerz sein. Dies setzt nach dem eben Gesagten voraus, dass sich die verschiedenen physischen Eigenschaften, die Schmerz (in verschiedenen Lebewesen oder verschiedenen Situationen)

[105] Vgl. Funkhouser (2006: 550).

[106] Funkhouser (2006: 563) diskutiert ein entsprechendes Beispiel für Überzeugungen. Überzeugungen können sich in ihrem Überzeugungsgrad und ihrem Gehalt unterscheiden. Determinaten der Determinablen ‚zu glauben, dass p' könnten entsprechend Eigenschaften wie ‚mit einem Überzeugungsgrad von 95 % zu glauben, dass Paris in Frankreich ist' sein.

zugrunde liegen, ebenfalls anhand von Unterscheidungsdimensionen unterscheiden lassen, die einen Bezug zu der Art und Weise haben, Schmerz zu sein. Es ist jedoch keinesfalls klar, dass die verschiedenen physischen Basen von Schmerz mit verschiedenen Arten und Weisen, Schmerz zu sein, einhergehen.

Vielmehr erlaubt die ursprüngliche Formulierung der These der multiplen Realisierung auf den ersten Blick, dass verschiedene Objekte auf *genau dieselbe Art und Weise* Schmerz instantiieren und dennoch verschiedene physische Basen instantiieren. Susi und Peter können beide starke, stechende Kopfschmerzen haben und dennoch verschiedene physische Basen instantiieren. Auch maximal spezifische mentale Eigenschaften können multipel realisiert sein. In diesem Fall könnten sich die unterschiedlichen physischen Basen von Schmerz gerade nicht mit Blick auf ihre Art und Weise, Schmerz zu sein, unterscheiden.[107]

Natürlich ließen sich nun *physische* Unterscheidungsdimensionen für Schmerz stipulieren: verschiedene Arten, Schmerz zu haben, können dann nicht bloß anhand der mentalen Dimensionen ‚Intensität', ‚Lokalisation' und ‚phänomenale Qualität' unterschieden werden, sondern außerdem anhand einer physischen Unterscheidungsdimension, die sich aus Unterschieden der physischen Basen ergibt. Susi und Peter würden dann entgegen der oben impliziten Annahme *nicht* eine maximal spezifische mentale Eigenschaft teilen, weil sich Susis stechender starker Kopfschmerz von Peters stechendem starken Kopfschmerz entlang der physischen Unterscheidungsdimension unterscheidet.[108]

Auch dieser Ausweg würde jedoch nahelegen, dass sich entsprechende physische Unterscheidungsdimensionen angeben lassen und dass sich die verschiedenen Basen anhand der physischen Unterscheidungsdimensionen unterscheiden lassen. Eine gewisse Geordnetheit der Klasse der physischen Basen einer mentalen Eigenschaft ist hier also vorausgesetzt. Allgemein *ähneln* sich die Determinaten einer Determinablen, wobei sie sich zugleich anhand der Unterscheidungsdimensionen auf spezifizierbare Weise voneinander unterscheiden. Die These, dass mentale Eigenschaften Determinablen mit physischen Determinaten sind, würde also zumindest Versionen der These multipler Realisierung ausschließen, denen zufolge die Basen mentaler Eigenschaften auf physischer Ebene auf beliebige Weise heterogen sein können.

Es bleibt also unklar, ob die Relation der Determination alles leisten kann, was im Rahmen eines nicht-reduktiven Physikalismus von einer geeigneten Relation ontologischer Abhängigkeit verlangt ist. Zumindest aber handelt es sich um einen weiteren vielversprechenden Kandidaten.

---

[107] Vgl. Funkhouser (2006: 563–565) und Walter (2007: 229–230).

[108] Vgl. hierzu Wilson (2009).

2.5.5 Mikrobasierte Eigenschaften

Die bis hierher diskutierten Relationen ontologischer Abhängigkeit haben gemeinsam, dass sie Eigenschaften, die an einem Objekt instantiiert sind, zu anderen Eigenschaften in Beziehung setzen, die an *demselben* Objekt instantiiert sind: Die funktionale Eigenschaft zweiter Ordnung, eine Eigenschaft zu haben, die die Schmerz Rolle hat, ist am selben Objekt instantiiert wie die realisierende Eigenschaft erster Ordnung, die die Schmerz-Rolle hat. Determinablen werden durch dieselben Objekte instantiiert wie ihre Determinaten. Und auch die oben angeführte Relation starker Supervenienz setzt voraus, dass mentale Eigenschaften an denselben Objekten instantiiert sind wie die physischen Eigenschaften in der Supervenienzbasis.

Realisierung und Determination setzen zudem eine Eigenschaft zu *einer* anderen Eigenschaft in Beziehung. Susis Schmerz kann zwar mal durch $\mathbf{N_1}$ und mal durch $\mathbf{N_2}$ realisiert sein. In beiden Fällen steht jedoch nur *eine* Eigenschaft zu Schmerz in der Realisierungsrelation. Die grundlegenden Realisierungsaussagen sind ‚$\mathbf{N_1}$ realisiert Schmerz' und ‚$\mathbf{N_2}$ realisiert Schmerz', aber nicht ‚$\mathbf{N_1}$ und $\mathbf{N_2}$ realisieren gemeinsam Schmerz'. Entsprechendes gilt für Determination: scharlachrot und bordeauxrot determinieren jeweils rot. Sie determinieren aber nicht gemeinsam rot. In diesem Sinne sind beide Relationen eins-zu-eins Relationen.[109]

Es finden sich hingegen auch Kandidaten für eine geeignete Relation ontologischer Abhängigkeit, die die mentalen und spezialwissenschaftlichen Eigenschaften eines Objekts zu mehreren physischen Eigenschaften *der Teile* des Objekts in Beziehung setzen. Diese ontologischen Abhängigkeitsrelationen setzen also Eigenschaften eines Objekts zu mehreren Eigenschaften anderer Objekte in Beziehung. Es handelt sich um eins-zu-viele Relationen. Mithilfe solcher Relationen kann der Idee Ausdruck verliehen werden, dass die mentalen und spezialwissenschaftlichen Eigenschaften makroskopischer Objekte ontologisch abhängen von den physischen Eigenschaften der mikroskopischen Teile, aus denen die makroskopischen Objekte zusammengesetzt sind.

Gerade bei recht restriktiven Bestimmungen der Klasse physischer Eigenschaften, die sich etwa an mikrophysikalischen Theorien orientieren, scheint eine Berufung auf solche Mikro/Makro-Relationen ontologischer Abhängigkeit unabdingbar. Denn unter dieser Voraussetzung finden sich möglicherweise schon gar keine ‚physischen' Eigenschaften (abgesehen vielleicht von Masse oder Ladung), die überhaupt von den Trägern mentaler Eigenschaften instantiiert werden. Die paradigmatischen Träger physischer Eigenschaften sind dann schließlich

[109] Vgl. auch Bennett (2017: 11).

mikroskopisch, während die paradigmatischen Träger mentaler Eigenschaften makroskopisch sind.

Eine solche eins-zu-viele Relation ontologischer Abhängigkeit ist Mikro-Basiertheit. Mikrobasierte Eigenschaften sind Eigenschaften zusammengesetzter Objekte, die durch ihre Mikrostruktur charakterisiert sind. Ein Diamant hat beispielsweise die mikrobasierte Eigenschaft, vollständig aus Kohlenstoffatomen zusammengesetzt zu sein, die in bestimmten Bindungsrelationen zueinander stehen. Allgemein können mikrobasierte Eigenschaften wie folgt charakterisiert werden:

> **Mikrobasierte Eigenschaft**: **F** ist genau dann eine mikrobasierte Eigenschaft, wenn gilt: **F** ist die Eigenschaft, vollständig aus distinkten Teilen a_1, a_2, …, a_n zusammengesetzt zu sein, so dass a_1 die Eigenschaft $\mathbf{F_1}$ hat, a_2 die Eigenschaft $\mathbf{F_2}$ hat, … und a_n die Eigenschaft $\mathbf{F_n}$ hat und $a_1 - a_n$ in Relation **R** zueinander stehen.[110]

Da die mikrobasierten Eigenschaften eines makroskopischen Objekts über die Eigenschaften und Relationen der Teile des Objekts definiert sind, gibt es einen klaren Sinn, in dem mikrobasierte Eigenschaften nicht mehr als die Eigenschaften und Relationen der Teile sind, auf denen sie basieren. Auch sind mikrobasierte Eigenschaften klarerweise *aufgrund* der Eigenschaften instantiiert, auf denen sie basieren. Der Diamant hat die Eigenschaft, aus Kohlenstoffatomen zusammengesetzt zu sein, die in bestimmten Bindungsrelationen zueinander stehen, *weil* seine Teile die Eigenschaft haben, Kohlenstoffatome zu sein und in den entsprechenden Bindungsrelationen zueinander stehen. Die mikrobasierten Eigenschaften makroskopischer Objekte sind also in einem recht starken Sinne ontologisch abhängig von den (mikroskopischen) Eigenschaften, auf denen sie basieren.

Obzwar die Verbindung zwischen mikrobasierten Eigenschaften und den Mikro-Eigenschaften, auf denen sie basieren, also sehr eng ist, scheinen die Aussichten, mentale und spezialwissenschaftliche Eigenschaften als mikrobasierte Eigenschaften aufzufassen, eher schlecht. Schon paradigmatische Makro-Eigenschaften wie Härte oder Masse sind plausiblerweise *nicht* mit mikrobasierten Eigenschaften identisch.[111] Denn auch für solche Eigenschaften scheint ein Pendant der These der multiplen Realisierung plausibel: Mein Schreibtisch und

[110] Vgl. Kim (1998: 84).

[111] Eine andere Frage ist es, ob mikrobasierte Eigenschaften in kausale Konkurrenz mit den Mikroeigenschaften treten, auf denen sie basieren. Kim (1998: 80–87; 112–118) meint, dass mikrobasierte Eigenschaften nicht von kausaler Exklusion bedroht sind (vgl. auch Kim

meine Nichte wiegen beide 20 kg. Es scheint jedoch wenig plausibel, dass sie dieselbe Zerlegung in Teile mit denselben Gewichtsanteilen haben.[112]

Dennoch kann der Begriff einer mikrobasierten Eigenschaft eine wichtige Funktion erfüllen: Mikrobasierte Eigenschaften erlauben nämlich, eine Brücke zu schlagen zwischen den in 2.5.3. und 2.5.4. diskutierten eins-zu-eins Relationen ontologischer Abhängigkeit und entsprechenden eins-zu-viele Behauptungen ontologischer Abhängigkeit. Man könnte zum Beispiel sagen, dass mikrobasierte Eigenschaften als Realisierer mentaler und spezialwissenschaftlicher Eigenschaften fungieren können. Die Füller der Schmerz-Rolle wären dann mikrobasierte Eigenschaften. Susis Schmerz wäre durch eine mikrobasierte Eigenschaft von Susi realisiert und der Schmerz des Oktopusses Paul durch eine andere mikrobasierte Eigenschaft von Paul. Da Schmerz dann durch die Realisierungsrelation von den entsprechenden mikrobasierten Eigenschaften ontologisch abhängt und die mikrobasierten Eigenschaften durch die Relation der Mikrobasiertheit von Mikroeigenschaften ontologisch abhängt, kann man entsprechend sagen, dass Schmerz von Mikroeigenschaften ontologisch abhängt.[113]

2.5.6 Grounding

Ein weiterer Kandidat für die gesuchte Relation der ontologischen Abhängigkeit ist Grounding. Was ist Grounding?

(2005: 52–69)). Er glaubt, mithilfe mikrobasierter Eigenschaften – contra Block (2003) – eine Generalisierung des Exklusionsproblems verhindern zu können. Walter (2008a: 686–689) argumentiert hiergegen, dass sich auch für mikrobasierte Eigenschaften ein Exklusionsproblem ergibt.

[112] Vgl. Block (2003: 145–146) und Walter (2008a: 690). Beide Autoren schließen aus solchen Beispielen, dass mikrobasierte Eigenschaften multipel zusammengesetzt sein können und daher nicht mit genauen Mikrostrukturen identisch sein können. Gegeben Kims Definition von mikrobasierten Eigenschaften scheint mir die korrekte Konsequenz hingegen zu sein, dass gewöhnliche Makro-Eigenschaften nicht mit mikrobasierten Eigenschaften identisch sind, da mikrobasierte Eigenschaften *per definitionem* mit genauen Mikrostrukturen identisch sind.

[113] Gillett (2002) und Gillett (2003) entwickelt zudem eine Variante der Realisierungsrelation, die zwischen Eigenschaften zusammengesetzter Objekte und den Eigenschaften ihrer Teile besteht. Auch Shoemaker (2007: 32–54) diskutiert eine Relation der Mikro-Realisierung. Auf diese Relationen gehe ich nicht mehr ein. Insofern es bloß darum geht, Behauptungen ontologischer Abhängigkeit zu rekonstruieren, die Makroeigenschaften zu Mikroeigenschaften in Beziehung setzen, scheint mir der Brückenschlag über mikrobasierte Eigenschaften hinreichend.

Gelegentlich wird ‚grounding' schlicht als ein generischer Begriff verwendet, unter den unterschiedliche Relationen ontologischer Abhängigkeit – wie die bereits diskutierten Relationen der Realisierung und der Determination – fallen.[114] So verstanden ist es unkontrovers, dass nicht-reduktive Physikalist*innen darauf festgelegt sind, dass mentale und spezialwissenschaftliche Eigenschaften in physischen Eigenschaften gegroundet sind. Es handelt sich bei der These, dass mentale und spezialwissenschaftliche Eigenschaften in physischen Eigenschaften gegroundet sind, dann schlicht um eine Reformulierung der These der ontologischen Abhängigkeit.

Einige Autor*innen verbinden jedoch weitaus mehr mit der Grounding-Relation. Jonathan Schaffer etwa spricht Grounding als „the primitive structuring conception of metaphysics"[115] an. Hierhinter stecken die Ideen, dass es erstens eine *einheitliche* Relation ontologischer Abhängigkeit – Grounding – gibt, die allen philosophischen Behauptungen über ontologische Abhängigkeit, ontologische Priorität, Fundamentalität und ontologische 'aufgrund von'-Beziehungen zugrunde liegt und dass diese Relation zweitens in dem Sinne *primitiv* ist, dass sie nicht unter Verwendung grundlegenderer Begriffe definiert werden kann. Kurzum: Grounding ist die einheitliche und primitive Relation ontologischer Abhängigkeit.[116]

Obgleich Grounding primitiv ist und daher nicht durch grundlegendere Begriffe definiert werden kann, kann man durchaus bestimmte Merkmale der Grounding-Relation herausarbeiten und interessante Verbindungen zwischen Grounding und anderen Begriffen aufzeigen. Zu diesem Projekt ist in den vergangenen Jahren eine ausführliche und verästelte philosophische Debatte entstanden, auf die ich hier nur in den gröbsten Grundzügen eingehen kann.

Zunächst zu den Relata der Grounding-Relation: Grounding wird oft als eine Relation ausschließlich zwischen *Tatsachen* bestimmt.[117] Hieraus ergibt sich die

---

[114] Wilson (2014: 539) und Wilson (2016: 496) redet in diesem Zusammenhang von 'small-g grounding relations' und unterscheidet dies von 'big-G Grounding'.

[115] Schaffer (2009: 364). Vgl. ähnlich auch Rosen (2010: 113–114).

[116] Beide Behauptungen werden natürlich nicht universal akzeptiert. Vgl. z. B. Daly (2012) und Koslicki (2015: 310–320) für Kritik an der Einheitlichkeitsthese und Correia (2013) für Diskussionen verschiedener Definitionen von Grounding über den Begriff der Essenz.

Ich konzentriere mich hier auf Grounding als einheitliche und primitive Relation ontologischer Abhängigkeit, weil so verstandenes Grounding am ehesten einen interessanten eigenständigen Kandidaten für die gesuchte Relation ontologischer Abhängigkeit darstellt. Zudem sind dies typische Thesen von 'Grounding-Enthusiast*innen' oder 'Groundhogs' (vgl. Koslicki (2020: 164) für diese Ausdrücke).

[117] Ich übergehe hier die Option, dass Grounding primär eine Beziehung zwischen Sätzen ist oder als Satzoperator aussgedrückt werden kann (vgl. z. B. Fine (2012: 46). Denn, wie Morris

Frage, wie die für den nicht-reduktiven Physikalismus ausschlaggebende Relation ontologischer Abhängigkeit, deren Relata *Eigenschaften* sind, als Grounding verstanden werden kann. Unter Voraussetzung einer starken Einheitlichkeitsthese, der zufolge Grounding wortwörtlich die *einzige* Relation ontologischer Abhängigkeit ist, würde zudem folgen, dass *ausschließlich* Tatsachen in ontologischen Abhängigkeitsbeziehungen stehen. Dies würde die Rekonstruktion vieler verbreiteter Thesen ontologischer Abhängigkeit erschweren.[118] Im vorliegenden Kontext ist es daher sinnvoll, eine neutrale Einstellung gegenüber den Relata der Grounding-Relation einzunehmen. Grounding kann demnach Entitäten unterschiedlichster Art miteinander verbinden. Insbesondere können auch Eigenschaften andere Eigenschaften grounden.[119]

Mit Blick auf ihre formalen Merkmale wird die Grounding-Relation in der Regel wie folgt charakterisiert: Grounding ist asymmetrisch, irreflexiv und transitiv. Wenn A B groundet, dann groundet B nicht A. Nichts groundet sich selbst. Und wenn A B groundet und B C groundet, dann groundet A auch C.[120] Zudem ist Grounding laut vielen Autor*innen hyperintensional:[121] Denn die Grounding-Relation kann asymmetrisch zwischen Tatsachen bestehen, die einander necessitieren. Kit Fine gibt hierfür das folgende Beispiel:[122] Sokrates groundet die Menge {Sokrates}, deren einziges Element Sokrates ist. Die Menge {Sokrates} existiert, weil Sokrates existiert. Aber Sokrates existiert nicht deshalb, *weil* die Menge {Sokrates} existiert. Da Sokrates und die Menge {Sokrates} in exakt denselben möglichen Welten existieren, besteht die Grounding-Relation in diesem Falle also zwischen zwei Tatsachen, die einander necessitieren. Entsprechend bleibt der Wahrheitswert der Aussage ‘Sokrates groundet {Sokrates}’ bei Austausch von ko-intensionalen[123] Ausdrücken nicht erhalten. Denn wenn der Ausdruck ‘Sokrates’ durch den ko-intensionalen Ausdruck {Sokrates} ersetzt wird, ergibt sich die falsche Aussage ‘{Sokrates} groundet {Sokrates}’. In

(2019: 161) betont: “for Grounding to play a substantive role in nonreductive physicalism, it will have to relate worldly items, as only in this case will it have a chance of articulating the contention that there are higher-level objects, properties, and processes, each and all of them dependent on and determined by physical items”.

[118] Vgl. Bennett (2017: 13).

[119] Schaffer (2009) setzt eine solche neutrale Einstellung gegenüber den Relata der Grounding-Relation voraus. Vgl. auch Raven (2020a: 6).

[120] Vgl. Thompson (2020) für kritische Diskussion.

[121] Vgl. Duncan et. al. (2017) für Diskussion.

[122] Vgl. Fine (1994) und Schaffer (2009: 364).

[123] Vorausgesetzt ist hier natürlich, dass Intensionen über mögliche Welten verstanden werden, und nicht etwa über inferentielle Rollen o.ä.

diesem Sinne ist Grounding hyperintensional. Außerdem ist die Grounding-Relation faktiv: daraus, dass Sokrates Existenz {Sokrates} Existenz groundet, folgt, dass sowohl Sokrates als auch {Sokrates} tatsächlich existieren. Schließlich ist Grounding nicht-monoton: Daraus, dass Sokrates Existenz die Existenz von {Sokrates} groundet, folgt nicht, dass Sokrates Existenz und Xanthippes Existenz die Existenz von {Sokrates} grounden.

Man kann partielles Grounding von vollständigem Grounding unterscheiden. Besonders einsichtig wird dies im Fall von konjunktiven Tatsachen: Der Ball ist rot und rund. Diese Tatsache ist teilweise in der Tatsache gegroundet, dass der Ball rot ist. Diese Tatsache groundet die konjunktive Tatsache jedoch nur partiell. Erst gemeinsam mit der Tatsache, dass der Ball rund ist, groundet sie die konjunktive Tatsache vollständig. Vollständiges Grounding kann entsprechend sowohl als eins-zu-eins-Relation als auch als eins-zu-viele Relation in Erscheinung treten. Es ist durchaus möglich und üblich, dass mehrere Tatsachen und Eigenschaften erst im Zusammenspiel eine andere Tatsache oder Eigenschaft vollständig grounden.[124] Daher spricht von dieser Seite auch nichts dagegen, die Beziehung zwischen den physischen Mikro-Eigenschaften eines Objekts und seinen Makro-Eigenschaften (inklusive mentaler Eigenschaften) als eine Grounding-Beziehung aufzufassen. Die Mikroeigenschaften der Teile des Objekts könnten die Makro-Eigenschaften jeweils partiell und erst im Zusammenspiel vollständig grounden. Mit Grounding geht – anders als im Fall von Realisierung und Determination – weder unbedingt einher, dass die Relation zwischen nur *zwei* Eigenschaften besteht noch, dass die Eigenschaften, die in der Relation stehen, am selben Objekt instantiiert sind.

Die meisten Autor*innen gehen von der folgenden Verbindung zwischen vollständigem Grounding und Necessitation aus: Dass A B metaphysisch necessitiert, ist notwendig, aber nicht hinreichend dafür, dass A B groundet. Dass Necessitation nicht hinreichend für Grounding ist, wird schon dadurch klar, dass Necessitation nicht asymmetrisch und irreflexiv ist. Jede Tatsache necessitiert sich selbst, aber keine Tatsache groundet sich selbst. Dass Necessitation notwendig für vollständiges Grounding ist, kann durch die Verbindung zwischen Grounding und der 'aufgrund von'-Redeweise motiviert werden. Wenn eine Tatsache A eine andere Tatsache B nicht necessitiert, scheint B eben *nicht* alleine aufgrund von B zu bestehen.[125]

---

[124] Vgl. Krämer und Schnieder (2017: 280).

[125] Diese knappe Motivation ist natürlich nicht ganz unproblematisch. Vgl. Skiles (2020: 150–151) für weitere Diskussion.

Dieser kurze Überblick muss als Charakterisierung der Grounding-Relation reichen. Es stellt sich nun die Frage, ob Grounding für eine Explikation der für den nicht-reduktiven Physikalismus ausschlaggebenden These der ontologischen Abhängigkeit fruchtbar gemacht werden kann. Ermöglicht die These, dass mentale und spezialwissenschaftliche Eigenschaften in physischen Eigenschaften gegroundet sind, also eine überzeugende Erklärung dafür, dass mentale und spezialwissenschaftliche Eigenschaften auf physischen Eigenschaften supervenieren und begründet diese These, dass mentale und spezialwissenschaftliche Eigenschaften *aufgrund von* physischen Eigenschaften instantiiert sind?

Insofern Necessitation wie oben geschildert notwendig für vollständiges Grounding ist, erklärt die These, dass mentale und spezialwissenschaftliche Eigenschaften durch physische Eigenschaften gegroundet sind, dass sie auf physischen Eigenschaften supervenieren. Zwei Objekte, die dieselben physischen Eigenschaften haben, instantiieren dann auch dieselben 'Grounds' für mentale und spezialwissenschaftliche Eigenschaften und somit notwendigerweise auch dieselben mentalen und spezialwissenschaftlichen Eigenschaften. Es ist jedoch nicht klar, dass es eine zufriedenstellende Erklärung dafür gibt, *warum* 'Grounds' die zugehörigen gegroundeten Eigenschaften necessitieren.[126]

Ähnlich verhält es sich mit der These, dass mentale und spezialwissenschaftliche Eigenschaften *aufgrund von* physischen Eigenschaften instantiiert sind. Grounding ist *per definitionem* mit 'aufgrund von'-Behauptungen verbunden. Insofern folgt aus der Behauptung, dass mentale und spezialwissenschaftliche Eigenschaften in physischen Eigenschaften gegroundet sind, auch, dass sie aufgrund von physischen Eigenschaften instantiiert sind. Es bleibt jedoch die Frage nach einer tieferliegende Erklärung für diese Verbindung. Warum grounden physische Eigenschaften mentale Eigenschaften?[127]

Nach meiner Einschätzung können die Relationen der Realisierung, der Determination oder der Mikrobasiertheit hier durchaus mehr leisten: Wenn Schmerz beispielsweise eine funktionale Eigenschaft ist, die durch die physische Eigenschaften $\mathbf{N_1}$ realisiert wird, ist die Relevanz der physischen Eigenschaft $\mathbf{N_1}$ für die Instantiierung von Schmerz unmittelbar einsichtig. Man erkennt, *was* $\mathbf{N_1}$ genau zur Instantiierung von Schmerz beiträgt: Denn $\mathbf{N_1}$ spielt die benötigte kausale Rolle. Eine entsprechende Erklärung lässt sich aus der bloßen Behauptung, dass $\mathbf{N_1}$ Schmerz (partiell oder vollständig) groundet, jedoch nicht ableiten. Man stellt sich dann vielmehr die Frage, warum denn gerade $\mathbf{N_1}$ dazu in

[126] Vgl. auch Melnyk (2016: 252–255).

[127] Vgl. Koslicki (2015: 309) und Bryant (2020: 493–494). Ich danke Jonas Werner für eine hilfreiche Anmerkung zu diesem Punkt.

der Lage ist, Schmerz zu grounden.[128] Zu einer primitiven Grounding-Relation, die die gewünschte Verknüpfung zwischen mentalen/spezialwissenschaftlichen Eigenschaften und physischen Eigenschaften letztlich per Stipulation etabliert, sollten nicht-reduktive Physikalist*innen nach meiner Einschätzung vorerst nur dann Zuflucht nehmen, wenn die anderen Kandidaten für eine Bestimmung der gesuchten Relation der ontologischen Abhängigkeit scheitern.

Auf der anderen Seite kann die relative Unbestimmtheit der Grounding-Relation auch als Vorteil gesehen werden. Der nicht-reduktive Physikalismus kann bei Berufung auf Grounding – anders als bei einer Berufung auf Realisierung – neutral gegenüber der Frage bleiben, ob mentale Zustände ihrer Natur nach funktional sind. Anders als bei einer Berufung auf Determination ist er nicht auf ein bestimmtes Verständnis der These der multiplen Realisierung festgelegt.

Zudem ist es durchaus möglich, dass die Annahme einer einheitlichen und primitiven Grounding-Relation sich in vielen unterschiedlichen philosophischen Kontexten als fruchtbar erweist.[129] Wenn es also – unabhängig von der spezifischen Frage nach der Formulierung des nicht-reduktiven Physikalismus – sinnvoll ist, eine primitive Grounding-Relation einzuführen, ist es sicherlich naheliegend, diese auch auf die Formulierung des nicht-reduktiven Physikalismus anzuwenden.[130]

---

[128] Einige Autor*innen geben zudem zu Bedenken, ob die These, dass mentale und spezialwissenschaftliche Eigenschaften in physischen Eigenschaften gegroundet sind, etabliert, dass mentale und spezialwissenschaftliche Eigenschaften im für den Physikalismus ausschlaggebenden Sinne ‘nicht mehr als’ physische Eigenschaften sind. So argumentiert etwa Wilson (2016: 503–507), dass die skizzenhaft angegebenen Merkmale der Grounding-Relation keinesfalls deutlich werden lassen, worin sich Grounding von einer starken Emergenzrelation unterscheidet, wie sie von erklärten (nicht-physikalistischen) Dualisten angeführt wird. Vgl. auch Melnyk (2016: 252–257) für eine ähnliche Argumentation.

[129] Vgl. Part V in Raven (2020b) für verschiedene philosophische Anwendungen der Grounding-Relation.

[130] Kroedel und Schulz (2016: 1921) sehen zudem konkrete Vorteile der Grounding-Relation in der Anwendung auf den nicht-reduktiven Physikalismus im Zusammenhang mit dem Vorwurf des Epiphänomenalismus. Soweit ich sehe sind dies jedoch Vorteile gegenüber einer bloßen Supervenienz-Version, nicht aber gegenüber einer Realisierungs- oder Determinationsversion des nicht-reduktiven Physikalismus.

Ney (2016) und (2020) zieht zudem die Möglichkeit in Betracht, einen nicht-skeptischen, antirealistischen Physikalismus mithilfe der Grounding-Relation zu formulieren. Soweit ich sehe kann eine solche Position jedoch auch ohne die Grounding-Relation formuliert werden – schlicht durch eine Ablehnung des realistischen Elements des nicht-reduktiven Physikalismus. Zugleich zeigt dies aber – vielleicht contra Wilson (2014: 540) – dass die Grounding-Relation, um für einen nicht-reduktiven Physikalismus fruchtbar gemacht zu werden, für sich genommen nicht zur Folge haben muss, dass die gegroundeten Eigenschaften

2.5.7 Zwischenbilanz: Ontologische Abhängigkeit im nicht-reduktiven Physikalismus

Die Diskussion in den vergangenen Abschnitten hat gezeigt, dass nicht-reduktiven Physikalist*innen mehrere Optionen offenstehen, die These der ontologischen Abhängigkeit zu präzisieren. Zwar leistet eine bloße Berufung auf Supervenienz für sich genommen keine Erklärung für den modalen Zusammenhang zwischen physischen und mentalen/spezialwissenschaftlichen Eigenschaften. Und auch lässt sich aus dem Bestehen einer Supervenienz-Beziehung nicht schließen, dass mentale und spezialwissenschaftliche Eigenschaften *aufgrund von* physischen Eigenschaften instantiiert sind. Daher sollte eine ausgearbeitete Version des nicht-reduktiven Physikalismus auf eine der gehaltvolleren Relationen Realisierung, Determination oder Grounding berufen.

Allerdings folgt hieraus nicht, dass es für die Zwecke dieser Arbeit notwendig ist, sich für eine dieser Optionen – Realisierung, Determination oder Grounding – zu entscheiden. Aus zwei Gründen ist eine neutrale Einstellung gegenüber der genauen Relation ontologischer Abhängigkeit im vorliegenden Kontext sogar vorteilhaft:

Erstens geht es mir in diesem Kapitel in erster Linie um eine neutrale Charakterisierung und Abgrenzung des nicht-reduktiven Physikalismus, nicht um eine detaillierte Entwicklung einer theoretisch einwandfreien Version des nicht-reduktiven Physikalismus. Für die detaillierte Entwicklung einer Version des nicht-reduktiven Physikalismus wäre eine Präzisierung der Relation ontologischer Abhängigkeit unabdingbar. Für eine Abgrenzung von anderen Positionen ist dies aber nicht nötig. Vielmehr reicht hierfür eine Festlegung darauf, dass mentale und spezialwissenschaftliche Eigenschaften von physischen Eigenschaften metaphysisch necessitiert werden.[131]

Zweitens meine ich, dass die Festlegung auf eine bestimmte Relation ontologischer Abhängigkeit auch für die Hauptfragestellungen in den weiteren Teilen dieser Arbeit keine große Relevanz hat. Es ist für keine der genannten Relationen offenbar, dass sie die für den in Teil 3 der Arbeit diskutierten Vorwurf des Epiphänomenalismus ausschlaggebenden kausalen Konkurrenzintuitionen *nicht* entstehen lässt. Exklusionsprobleme ergeben sich daher prima facie bei jeder

real, genuin oder spärlich sind. Dies kann vielmehr als zusätzliche These explizit benannt werden.

[131] In Abschnitt 2.6.6. qualifiziere und begründe ich diesen – nicht ganz unkontroversen – Punkt noch einmal ausführlicher.

der genannten Relationen.[132] Und für die in Teil 4 der Arbeit diskutierten Erwiderungen auf den Vorwurf des Epiphänomenalismus, die auf verschiedenen Theorien der Kausalität beruhen, scheinen mir ebenfalls in erster Linie die modalen Verknüpfungen zwischen physischen und mentalen/spezialwissenschaftlichen Eigenschaften relevant zu sein. Die verschiedenen Ansätze zur Erklärung dieser modalen Verknüpfungen machen hingegen keinen Unterschied.

Aus meiner Sicht kann der Teil der These der ontologischen Abhängigkeit, der für eine Abgrenzung des nicht-reduktiven Physikalismus und für die weitere Argumentation der Arbeit relevant ist, also in rein modalen Begrifflichkeiten spezifiziert werden. Wichtig ist zunächst eine These der universalen Necessitation:

> **Universale Necessitation**: Für jede instantiierte mentale/spezialwissenschaftliche Eigenschaft M gibt es eine instantiierte physische$_{\text{eng}}$ Eigenschaft P_t, so dass metaphysisch notwendig gilt: Wenn P_t instantiiert ist, dann ist M instantiiert.[133]

Diese These folgt nach jeder der oben besprochenen Kandidaten für die Relation der ontologischen Abhängigkeit aus der These der ontologischen Abhängigkeit. Eine physische Eigenschaft P_t, die eine mentale Eigenschaft M in diesem Sinne necessitiert, nenne ich – in Anlehnung an die in 2.5.3. diskutierte Unterscheidung zwischen totaler Realisierung und Kern-Realisierung und die in 2.5.6. diskutierte Unterscheidung zwischen partiellem und vollständigem Grounding – eine *totale* physische *Basis* für M.[134] Totale Basen für mentale und spezialwissenschaftliche Eigenschaften werden plausiblerweise recht komplexe konjunktive oder mikrobasierte[135] physische Eigenschaften sein, in denen

---

[132] In Abschnitt 3.4.3. gehe ich noch einmal kurz auf Konkurrenzintuitionen im Zusammenhang mit Determination ein.

[133] Vgl. Stoljar (2010: 109–126) für eine Formulierung des Physikalismus, die sich auf eine solche These der universalen Necessitation beruft.

[134] Vielleicht wäre es sinnvoll, hier noch eine Minimalitätsforderung zu ergänzen, etwa so: *Pt* ist genau dann eine minimale totale Basis für *M*, wenn kein ‘Bestandteil’ *B* von *Pt* eine totale Basis für *M* ist.

[135] Hierbei müsste natürlich vorausgesetzt sein, dass die Klasse physischer$_{\text{eng}}$ Eigenschaften unter Mikrobasiertheit geschlossen ist, d. h. dass mikrobasierte Eigenschaften, die ausschließlich auf physischen$_{\text{eng}}$ Eigenschaften basieren, ebenfalls als physisch$_{\text{eng}}$ zählen. Dies hätte den Vorteil, dass sich auch bei sehr engen Bestimmungen physischer Eigenschaften keine Frage stellt, ob die paradigmatischen Träger mentaler und spezialwissenschaftlicher Eigenschaften überhaupt (geeignete) physische$_{\text{eng}}$ Eigenschaften instantiieren. Auf der anderen Seite wird durch diese Festsetzung zumindest die Möglichkeit eröffnet, dass sich ein

möglicherweise auch ein Bezug auf physische Gesetze kodiert ist. Diese necessitieren mentale/spezialwissenschaftliche Eigenschaften im metaphysischen Sinn von Notwendigkeit. Totale Basen lassen sich von *Kernbasen* unterscheiden. Kernbasen sind dabei (in irgendeinem Sinne) Bestandteile totaler Basen, die (in irgendeinem Sinne) besonders salient[136] für die Instantiierungen der zugehörigen mentalen/spezialwissenschaftlichen Eigenschaft sind. Sie necessitieren die zugehörige mentale/spezialwissenschaftliche Eigenschaft nicht für sich genommen, sondern nur im Verbund mit geeigneten Umständen.

Wie ich meine, ist die These der universalen Necessitation als Konsequenz der These der ontologischen Abhängigkeit hinreichend für eine neutrale und abgrenzende Charakterisierung des nicht-reduktiven Physikalismus und für die in Teil 4 skizzierten Verteidigungsstrategien gegen den Vorwurf des Epiphänomenalismus. Diese Auffassung ist damit kompatibel, dass nicht-reduktive Physikalist*innen die These der universalen Necessitation nicht unerklärt stehenlassen sollten. Eine Erklärung der These der universalen Necessitation durch eine gehaltvollere Ausdeutung der These der ontologischen Abhängigkeit ist jedoch nicht das Projekt dieser Arbeit. Meine eigene Einschätzung ist, dass die These der universalen Necessitation auf verschiedene Weisen auf eine physikalistisch akzeptable Art erklärt werden kann und so verschiedene Versionen des nicht-reduktiven Physikalismus entstehen. Zugleich sehe ich durchaus, dass es ein Problem für den nicht-reduktiven Physikalismus wäre, wenn keine Erklärung für die These der universalen Necessitation gegeben werden könnte.[137]

In Anbetracht dieser Beobachtungen fahre ich daher damit fort, in einem neutralen und unspezifischen Sinne von ontologischer Abhängigkeit zu sprechen. Minimal ist mit dieser Rede von ontologischer Abhängigkeit dabei immer eine entsprechende Necessitationsbehauptung verbunden. Zudem gehe ich davon aus, dass mit ontologischer Abhängigkeit eine – wie auch immer im Detail geartete – Erklärung des angesprochenen modalen Zusammenhangs einhergeht und dass auf Grundlage von Thesen ontologischer Abhängigkeit entsprechende ‚aufgrund von'-Behauptungen begründet werden können.

Exklusionsproblem innerhalb des Bereichs physischer$_{eng}$ Eigenschaften ergibt. Das wäre nämlich dann der Fall, wenn mikrobasierte Eigenschaften mit den Mikroeigenschaften, auf denen sie basieren, in kausale Konkurrenz treten.

[136] Im Falle einer *Realisierung* mentaler/spezialwissenschaftlicher Eigenschaften durch ihre Basen ist diese Salienzbedingung über das Kriterium des Füllers der kausalen Rolle einigermaßen klar. Im Falle einer Determinations- oder Grounding-Relation ist die Sache weniger klar.

[137] Morris (2019) bietet eine umfassende Kritik verschiedener Kandidaten für die gesuchte erklärende Relation ontologischer Abhängigkeit.

Schließlich möchte ich im Folgenden noch knapp auf zwei weitere Themen eingehen, die üblicherweise mit dem nicht-reduktiven Physikalismus und der Relation der ontologischen Abhängigkeit in Verbindung gebracht werden: Im kommenden Abschnitt geht es um Fundamentalität, im darauffolgenden Abschnitt um das Stufenmodell der Realität.

2.5.8 Ontologische Abhängigkeit und absolute Fundamentalität

Hinter der Rede von Fundamentalität steckt die Idee, dass es einige Eigenschaften gibt, die in einem absoluten Sinne grundlegend oder basal sind. Diese Eigenschaften bilden die Grundlage für alle weiteren Eigenschaften. Sie sind die Eigenschaften, die Gott laut dem illustrativen Schöpfungsmythos aus Abschnitt 2.5.1. explizit erschaffen musste. Alle weiteren Eigenschaften ergeben sich dann aus den fundamentalen Eigenschaften. Unter der Voraussetzung, dass es solche absolut fundamentalen Eigenschaften gibt, kann auch der Physikalismus unter Bezug auf Fundamentalität bestimmt werden. Er wird dann zu der These, dass alle instantiierten *fundamentalen* Eigenschaften physisch sind.

Zwischen dieser Idee der Fundamentalität und der Relation der ontologischen Abhängigkeit besteht offenbar ein enger Zusammenhang. Tatsächlich ist es recht naheliegend, den Begriff der Fundamentalität unter Bezug auf ontologische Abhängigkeit zu definieren.[138] Dabei gibt es zwei eng verwandte Möglichkeiten: Erstens können die fundamentalen Eigenschaften als diejenigen Eigenschaften bestimmt werden, die von keinen anderen Eigenschaften ontologisch abhängen. Fundamentale Eigenschaften werden dann dadurch herausgegriffen, dass sie *nicht* in bestimmten ontologischen Abhängigkeitsbeziehungen stehen. In diesem Falle können wir von *Fundamentalität als Unabhängigkeit* sprechen. Zweitens können fundamentale Eigenschaften aber auch als diejenigen Eigenschaften bestimmt werden, von denen alle weiteren Eigenschaften ontologisch abhängen. In diesem Falle werden die fundamentalen Eigenschaften dadurch herausgegriffen, dass sie in bestimmten ontologischen Abhängigkeitsbeziehungen zu anderen Eigenschaften stehen. Fundamentale Eigenschaften werden hier also als Elemente einer minimal vollständigen Menge von Eigenschaften verstanden, aus

[138] Es gibt Autor*innen, die Fundamentalität nicht über ontologische Abhängigkeit verstehen. Lewis (1986a: 60) bestimmt den Begriff der Fundamentalität über den Begriff der Natürlichkeit. Wilson (2014: 560) hält den Begriff der Fundamentalität für primitiv. Bennett (2017: 129–136) argumentiert überzeugend gegen diese Alternativen.

deren Elementen sich alle weiteren Eigenschaften ergeben. Man kann daher von *Fundamentalität als Vollständigkeit* sprechen.[139]

Die beiden Bestimmungen führen zu Unterschieden, wenn die Relation der ontologischen Abhängigkeit einige der ihr üblicherweise zugeschriebenen Eigenschaften – insbesondere Irreflexivität, Transitivität oder Asymmetrie – *nicht* hat. Wenn ontologische Abhängigkeit beispielsweise nicht transitiv ist, kann es Eigenschaften geben, die von anderen Eigenschaften ontologisch abhängen und dennoch benötigt werden, damit sich weitere Eigenschaften ergeben. Schmerz könnte dann beispielsweise ausschließlich von der neuronalen Eigenschaft $\mathbf{N_1}$ abhängen und die neuronale Eigenschaft $\mathbf{N_1}$ von der physischen Eigenschaft **P**, ohne dass Schmerz von **P** abhängt. Eine minimal vollständige Menge müsste dann $\mathbf{N_1}$ enthalten, da Schmerz nur von $\mathbf{N_1}$ abhängt. Zugleich wäre $\mathbf{N_1}$ aber nicht unabhängig, weil sie von **P** abhängt.[140] Insofern ontologische Abhängigkeit aber irreflexiv, transitiv und asymmetrisch ist, laufen die Bestimmungen von Fundamentalität als Unabhängigkeit und die Bestimmung von Fundamentalität als Vollständigkeit auf dasselbe hinaus: Es zählen jeweils die genau selben Eigenschaften als fundamental.[141]

Aus der These der ontologischen Abhängigkeit – alle mentalen/spezialwissenschaftlichen Eigenschaften sind von physischen Eigenschaften ontologisch abhängig – folgt nach dieser Bestimmung unmittelbar, dass mentale und spezialwissenschaftliche Eigenschaften nicht fundamental sind. Ich nenne im Folgenden alle nicht-fundamentalen Eigenschaften *höherstufig*. Mentale und spezialwissenschaftliche Eigenschaften sind laut dem nicht-reduktiven Physikalismus also in diesem Sinne höherstufig: Sie hängen von anderen Eigenschaften ontologisch ab. Es ergibt sich so eine ausschließende und erschöpfende Unterscheidung zwischen fundamentalen, physischen Eigenschaften auf der einen Seite und höherstufigen Eigenschaften auf der anderen Seite: Jede Eigenschaft ist entweder fundamental oder höherstufig und keine Eigenschaft ist sowohl fundamental als auch höherstufig.

---

[139] Vgl. Bennett (2017: 105–111).

[140] Dies soll nur als Beispiel dienen, vgl. auch Bennett (2017: 113–114). Ob diese Situation mit dem nicht-reduktiven Physikalismus vereinbar wäre, ist eine andere Frage.

[141] Man könnte zunächst denken, dass Fundamentalität als Vollständigkeit fundamentale Eigenschaften verbietet, von denen keine weiteren Eigenschaften ontologisch abhängen. Solche Eigenschaften könnten dennoch unabhängig sein und daher als fundamental im Sinne von Fundamentalität als Unabhängigkeit gelten. Das ist jedoch nicht so: Insofern diese Eigenschaften unabhängig sind, ergibt sich ihre Instantiierung nicht aus der Instantiierung anderer Eigenschaften. Sie müssen daher in jeder minimal vollständigen Menge enthalten sein, weil sonst nichts für ihre eigene Instantiierung aufkäme. Vgl. Bennett (2017: 117).

Aus der These der ontologischen Abhängigkeit folgt jedoch nicht ohne weitere Annahmen, dass physische Eigenschaften fundamental sind. Dies hat zwei Gründe – und einer dieser Gründe macht eine Ergänzung zur These der ontologischen Abhängigkeit notwendig:

Erstens folgt aus der These der ontologischen Abhängigkeit nicht, dass überhaupt irgendwelche Eigenschaften absolut fundamental sind. Es wäre ja durchaus möglich, dass jede Eigenschaft von noch grundlegenderen Eigenschaften ontologisch abhängt. Es gäbe dann unendliche Ketten von immer grundlegenderen Eigenschaften, ohne dass aber je eine absolut fundamentale Eigenschaft zu finden ist.[142] Der Physikalismus scheint als solcher nicht auf die Existenz absolut fundamentaler Eigenschaften festgelegt zu sein.[143] Es ist also keine Schwäche der These der ontologischen Abhängigkeit, dass sie diese Möglichkeit offenlässt.

Zweitens aber folgt aus der These der ontologischen Abhängigkeit auch nicht, dass physische Eigenschaften ihrerseits nicht von noch grundlegenderen nicht-physischen Eigenschaften abhängen. Mentale und spezialwissenschaftliche Eigenschaften könnten beispielsweise von physischen Eigenschaften ontologisch abhängen, während alle physischen Eigenschaften von den nicht-physischen Eigenschaften Gottes ontologisch abhängen, die wiederum fundamental sind. In einem solchen Falle wäre es wohl plausibel, dass der Physikalismus falsch ist. Will man den Physikalismus dennoch nicht auf die These festnageln, dass physische Eigenschaften absolut fundamental sind, muss man als Ergänzung zumindest verlangen, dass physische Eigenschaften ausschließlich von physischen Eigenschaften ontologisch abhängen. Selbst bei einer unendlichen Kette ontologischer Abhängigkeiten würde dann ab einem bestimmten Punkt der Bereich des Physischen nicht mehr verlassen.

Trotz dieser Komplikation gehe ich an einigen Stellen vereinfachend davon aus, dass es eine absolut fundamentale Klasse von ausschließlich physischen Eigenschaften gibt. Soweit ich sehe würde ein Wechsel zu einem Modell, in dem es eine Klasse physischer Eigenschaften gibt, die ausschließlich von anderen physischen Eigenschaften abhängen, ohne dass es aber absolut fundamentale Eigenschaften gibt, keine großen Unterschiede machen und lediglich zu umständlicheren Formulierungen führen.

---

[142] Vgl. Schaffer (2003a).

[143] Vgl. auch Stoljar (2010: 35–36).

2.5.9 Ontologische Abhängigkeit und das Stufenmodell der Realität

Es ist eine Sache, zu behaupten, dass physische Eigenschaften fundamental sind und alle nicht-physischen Eigenschaften höherstufig sind. Eine andere Sache ist es, auch die höherstufigen Eigenschaften ihrerseits noch einmal in eine hierarchische Ordnung zu bringen. Typischerweise operieren nicht-reduktive Physikalist*innen nicht bloß mit einer Zweiteilung aller Eigenschaften in fundamentale und höherstufige Eigenschaften. Stattdessen gehen sie davon aus, dass die Realität in zahlreiche Stufen eingeteilt ist und auch die höherstufigen Eigenschaften dabei nicht alle derselben Stufe zuzuordnen sind. Hier kommt also ein *Stufenmodell der Realität* ins Spiel.[144]

Die Grundidee des Stufenmodells der Realität wird etwa von Ned Block passend auf den Punkt gebracht:

> The family of mental properties can be used to characterize the level of psychology, the family of neurological properties can be used to characterize the level of neuroscience, the family of elementary particle properties can be used to characterize the level of elementary particle physics. These distinct branches of science and their associated families of properties plausibly form a supervenience hierarchy […]. No mental difference without a neurological difference. No neurological difference without a bio-chemical difference. No bio-chemical difference without a difference in atomic physics. No difference in atomic physics without an elementary particle difference.[145]

Drei Ideen sind dabei ausschlaggebend:
Erstens lassen sich die Eigenschaften, die unterschiedlichen Stufen angehören, grob den verschiedenen Spezialwissenschaften zuordnen. Die Klasse der psychologischen Eigenschaften bildet also beispielsweise eine Stufe, die von der Stufe der neurologischen Eigenschaften unterschieden ist. Auch die Eigenschaften, die in der Physik behandelt werden, formen eine Stufe – die Stufe der $\text{physischen}_{\text{eng}}$ Eigenschaften.

Zweitens sind die Stufen durch eine Relation der ontologischen Abhängigkeit hierarchisch geordnet. Block geht im obigen Zitat von einer Ordnung durch die Supervenienz-Relation aus. Aber auch eine andere Relation der ontologischen Abhängigkeit – zum Beispiel Realisierung oder Determination – kann die strukturierende Funktion übernehmen. Die Idee ist also, dass mentale Eigenschaften

[144] Ein klassischer Bezugspunkt für das Stufenmodell der Realität ist Oppenheim und Putnam (1958). Im nicht-reduktiven Physikalismus sollte die Hierarchie natürlich – anders als bei Oppenheim und Putnam – nicht als eine Hierarchie der *Reduktion* verstanden werden.

[145] Block (2003: 142).

auf einer höheren Stufe stehen als neurologische Eigenschaften, weil mentale Eigenschaften von neurologischen Eigenschaften ontologisch abhängen. Neurologische Eigenschaften stehen hingegen auf einer höheren Stufe als $\text{physische}_{\text{eng}}$ Eigenschaften, weil neurologische Eigenschaften von $\text{physischen}_{\text{eng}}$ Eigenschaften ontologisch abhängen. Insofern $\text{physische}_{\text{eng}}$ Eigenschaften dann von keinen anderen Eigenschaften ontologisch abhängen, bilden sie die niedrigste, absolut fundamentale Stufe.[146]

Drittens sind die Eigenschaften, die einer Stufe angehören, nicht mit den Eigenschaften identisch, die einer anderen Stufe angehören. Mentale Eigenschaften sind beispielsweise nicht mit neurologischen Eigenschaften identisch, weil sie durch neurologische Eigenschaften multipel realisierbar sind und neurologische Eigenschaften sind nicht mit $\text{physischen}_{\text{eng}}$ Eigenschaften identisch, weil sie auf der Stufe $\text{physischer}_{\text{eng}}$ Eigenschaften multipel realisierbar sind. Die Beziehung zwischen den Eigenschaften verschiedener Stufen ist also durch ontologische Abhängigkeit und multiple Realisierbarkeit gekennzeichnet.[147]

---

[146] Diese Darstellung übergeht eine Komplikation: Bei Relationen ontologischer Abhängigkeit, die – anders als Supervenienz – nicht direkt ganze Klassen von Eigenschaften miteinander verbinden, stellt sich die Frage, wie Eigenschaften, die *nicht* in einer Relation ontologischer Abhängigkeit stehen, hinsichtlich ihrer Stufenzugehörigkeit verglichen werden können: Schmerz ist zwar klarerweise auf einer höheren Stufe als die neuronale Eigenschaft $\mathbf{N_1}$, von der Schmerz abhängt. Wieso aber gilt Schmerz als auf einer höheren Stufe stehend als die neuronale Eigenschaft $\mathbf{N_x}$, die in keiner Relation ontologischer Abhängigkeit zu Schmerz (oder irgendeiner anderen mentalen Eigenschaft) steht? Bennett (2017: 137–186) geht ausführlich darauf ein, wie die hier zugrundeliegende Relation relativer Fundamentalität unter Bezug auf ontologische Abhängigkeit definiert werden kann. Da es in den weiteren Teilen der Arbeit, soweit ich mich auf das Stufenmodell beziehe, aber immer um Eigenschaften geht, die in einer Relation ontologischer Abhängigkeit zueinander stehen, gehe ich auf diese Komplikation nicht näher ein. Wesentlich ist nur das folgende Konditional, das auch Bennett (2017: 40–41) für unkontrovers hält: Wenn eine Eigenschaft F von einer Eigenschaft G ontologisch abhängt, ist F weniger fundamental als G (und somit auf einer höheren Stufe als G).

[147] Am Stufenmodell der Realität lässt sich auch noch einmal deutlich machen, warum die Formulierung des (reduktiven und nicht-reduktiven) Physikalismus einen engen Sinn von physischen Eigenschaften voraussetzen sollte: Geht man von einem weiten Sinn physischer Eigenschaften aus, ist der reduktive Physikalismus mit einem Stufenmodell der Realität vereinbar (solange der Bereich mentaler Eigenschaften keine eigenständige Stufe bildet). Der reduktive Physikalismus wird aber üblicherweise so aufgefasst, dass er mit einem Stufenmodell der Realität nicht vereinbar ist. Zudem: Wäre der reduktive Physikalismus mit einem Stufenmodell der Realität vereinbar, würde er – insofern er ein solches akzeptiert – mit Blick auf das Problem der mentalen Verursachung gleichauf mit dem nicht-reduktiven Physikalismus sein.

Die These der ontologischen Abhängigkeit – mentale und spezialwissenschaftliche Eigenschaften sind von $\text{physischen}_{\text{eng}}$ Eigenschaften ontologisch abhängig – hat für sich genommen nicht ein komplexeres Stufenmodell der Realität zur Konsequenz. Zwar folgt aus der These der ontologischen Abhängigkeit, dass mentale und spezialwissenschaftliche Eigenschaften auf einer höheren Stufe stehen als $\text{physische}_{\text{eng}}$ Eigenschaften. Es folgt aus ihr jedoch zum Beispiel nicht, dass mentale Eigenschaften auf einer höheren Stufe stehen als neurologische Eigenschaften. Ich gehe davon aus, dass auch der nicht-reduktive Physikalismus – was seine physikalistischen Verpflichtungen anbelangt – nicht auf mehr festgelegt ist als auf die These der ontologischen Abhängigkeit.[148] Das Stufenmodell der Realität ist entsprechend keine notwendige Bedingung für den nicht-reduktiven Physikalismus. Jedoch wird der nicht-reduktive Physikalismus typischerweise mit einem Stufenmodell der Realität verbunden. Daher formuliere ich auch einige Punkte in Teil 3 und Teil 4 dieser Arbeit unter Bezug auf ein Stufenmodell der Realität.[149]

2.6 Verhältnisbestimmungen und Abgrenzungen

Im Folgenden soll der nicht-reduktive Physikalismus zu bestimmten anderen Thesen und Theorien in Verhältnis gesetzt werden, die häufig mit ihm in Verbindung gebracht werden. Dabei geht es auch um den Nachweis, dass der nicht-reduktive Physikalismus eine eigenständige Position ist, die von anderen Positionen klar abgegrenzt werden kann. Entgegen der Behauptung einiger Autor*innen[150] kollabiert der nicht-reduktive Physikalismus nicht unbedingt in eine andere Position.

Ich verfahre wie folgt: In Abschnitt 2.6.1. gehe ich auf das Verhältnis des nicht-reduktiven Physikalismus zu einer These der Ereignis-Identität ein, in Abschnitt 2.6.2. auf eine These der Token-Identität mit Bezug auf Eigenschaftsinstantiierungen. Abschnitt 2.6.3. geht auf das Verhältnis zum anomalen Monismus ein. Abschnitt 2.6.4. diskutiert das Verhältnis des nicht-reduktiven Physikalismus zu einer These, die ich mentalen Exzeptionalismus nenne. Abschnitt 2.6.5. geht auf die Abgrenzung des nicht-reduktiven Physikalismus von den physikalistischen

[148] Mit der in Abschnitt 2.5.8 erwähnten Ergänzung, dass physische Eigenschaften absolut fundamental sind oder nur von weiteren physischen Eigenschaften ontologisch abhängen.

[149] Vgl. insbesondere meine Ausführungen in 3.3.4. – 3.3.8.

[150] Vgl. klassisch Kim (1989: 40): „[I]f a relation is weak enough to be nonreductive, it tends to be too weak to serve as a dependence relation; conversely, when a relation is strong enough to give us dependence, it tends to be too strong – strong enough to imply reducibility“.

Alternativen des reduktiven Physikalismus und des eliminativen Physikalismus ein. Abschnitt 2.6.6. grenzt den nicht-reduktiven Physikalismus vom nomologischen Dualismus ab. In Abschnitt 2.6.7. erläutere ich vorläufig die Beziehung zwischen dem nicht-reduktiven Physikalismus und Thesen, die den *kausalen* Status mentaler, spezialwissenschaftlicher und physischer Ereignisse betreffen.

2.6.1 Ereignis-Identität

Zwei der für den nicht-reduktiven Physikalismus historisch wichtigsten Aufsätze sind sicherlich Donald Davidsons ‚Mental Events' und Jerry Fodors ‚Special Sciences, Or: The Disunity of Science as a Working Hypothesis'. In beiden Aufsätzen spielt eine Ereignis-Identitätsthese eine wichtige Rolle:[151]

> **Ereignis-Identität**: Mentale/spezialwissenschaftliche Ereignisse sind mit physischen Ereignissen identisch.

Gelegentlich finden sich in der Literatur – anknüpfend an Davidson und Fodor – Charakterisierungen des nicht-reduktiven Physikalismus, die dieser These eine wichtige Rolle zugestehen. Einige Autor*innen gehen dabei davon aus, dass bereits die Ereignis-Identitätsthese die physikalistischen Verpflichtungen des nicht-reduktiven Physikalismus hinreichend einfängt.[152] Tatsächlich ist das Verhältnis zwischen Ereignis-Identität und dem nicht-reduktiven Physikalismus jedoch komplexer.

Der Gehalt der Ereignis-Identitätsthese ist wesentlich davon abhängig, welche Ontologie von Ereignissen ihr zugrunde gelegt wird. Gehen wir von Kims Konzeption von Ereignissen aus, können nicht-reduktive Physikalist*innen die Ereignis-Identitätsthese nicht akzeptieren. Wie in 2.2.2. geschildert sind Ereignisse laut Kim Instantiierungen von Eigenschaften an Objekten zu Zeitpunkten oder Zeitintervallen. Die Ereignisse [o, E, t] und [o', E', t '] sind genau dann

[151] Vgl. Davidson (1970) und Fodor (1974: 100).

[152] Vgl. z. B. van Riel (2014a: 137): "Non-reductive physicalism, in its ontological version, consists in the affirmation of token-reduction and the denial of type-reduction for the relevant class of tokens and types." Vgl. auch Haug (2011: 455): "The standard way of drawing the distinction between reductive and nonreductive physicalism is to claim that […] nonreductive physicalism can be characterized as endorsing token-identity while denying type-identity". Haug weist diese Charakterisierung – teilweise aus den hier geschilderten Gründen – ebenfalls zurück.

identisch, wenn die konstitutiven Objekte o und o‘ identisch sind, die konstitutiven Zeitintervalle t und t‘ identisch sind *und die konstitutiven Eigenschaften E und E‘ identisch sind*. Wenn mentale Eigenschaften also nicht mit physischen Eigenschaften identisch sind, dann sind auch mentale Ereignisse – d. h. Ereignisse mit konstitutiven mentalen Eigenschaften – nicht mit physischen Ereignissen – d. h. Ereignisse mit konstitutiven physischen Eigenschaften – identisch. Unter Voraussetzung von Kims Konzeption von Ereignissen ist die Ereignis-Identitätsthese daher keine notwendige Bedingung für den nicht-reduktiven Physikalismus. Ganz im Gegenteil ist die *Ablehnung der Ereignis-Identitätsthese* dann eine notwendige Bedingung für den nicht-reduktiven Physikalismus. Denn eine so verstandene Ereignis-Identitätsthese wäre deutlich zu stark: Sie würde Eigenschaftsidentitäten voraussetzen, die nicht-reduktive Physikalist*innen nicht akzeptieren können.[153]

Gehen wir hingegen von Davidsons Konzeption von Ereignissen aus, ist die Ereignis-Identitätsthese deutlich zu schwach, um die physikalistischen Verpflichtungen des nicht-reduktiven Physikalismus einzufangen. Bei Davidson sind Ereignisse – wie in 2.2.2. geschildert – raumzeitlich individuierte Individuen, die verschiedene Eigenschaften[154] instantiieren. Mentale Ereignisse sind dadurch ausgezeichnet, dass sie mentale Eigenschaften instantiieren und physische Ereignisse sind dadurch ausgezeichnet, dass sie physische Eigenschaften instantiieren. Dasselbe Ereignis ist also sowohl mental als auch physisch, insofern es sowohl mentale als auch physische Eigenschaften instantiiert. Die Ereignisidentitätsthese läuft entsprechend darauf hinaus, dass jedes Ereignis, das eine mentale Eigenschaft instantiiert, auch irgendeine physische Eigenschaft instantiiert. Über das Verhältnis zwischen mentalen und physischen Eigenschaften wird hier also nicht mehr gesagt, als dass beide Arten von Eigenschaften von denselben Ereignissen instantiiert werden können. Das aber ist damit kompatibel, dass mentale Eigenschaften ansonsten gänzlich unabhängig von physischen Eigenschaften sind. Aus einer so verstandenen Ereignis-Identitätsthese folgt also keine These der ontologischen Abhängigkeit in Bezug auf Eigenschaften. Eine solche ist jedoch notwendig für den nicht-reduktiven Physikalismus. Also fängt die Ereignis-Identitätsthese unter Voraussetzung von Davidsons Ereigniskonzeption die physikalistischen Verpflichtungen des nicht-reduktiven Physikalismus nicht angemessen ein.[155]

Der nicht-reduktive Physikalismus ist also nicht ohne weitere Qualifikation auf eine Ereignis-Identitätsthese festgelegt. Unter der in den weiteren Teilen dieser

---

[153] Vgl. Haug (2011: 456).

[154] Auf Davidsons Einstellung zu Eigenschaften gehe ich in 2.6.2. noch einmal ein.

[155] Vgl. Haug (2011: 456–457).

Arbeit vorausgesetzten Ereigniskonzeption von Kim ist er im Gegenteil sogar auf eine Ablehnung der Ereignis-Identitätsthese festgelegt.[156]

2.6.2 Token-Identität und Eigenschaftsinstantiierungen

Die Ereignis-Identitätsthese wird häufig auch als ‚Token-Identitätsthese' eingeführt.[157] Die Voraussetzung dieser Gleichsetzung ist, dass Ereignisse als ‚Token' und nicht als ‚Typen' aufgefasst werden. Diese Voraussetzung scheint jedoch sowohl bei Kims als auch bei Davidsons Ereigniskonzeption gegeben zu sein: Schließlich sind Ereignisse in beiden Konzeptionen nicht wiederholbar. Es handelt sich also um konkrete Vorkommnisse (oder ‚Token') und nicht um Kategorien (oder ‚Typen'), unter die verschiedene konkrete Vorkommnisse fallen können. Bei Davidson liegt das an der raumzeitlichen Individuierung von Ereignissen, bei Kim an dem Verweis auf ein konstitutives Zeitintervall.

Die Ereignis-Identitätsthese ist jedoch nicht die einzige mögliche Entwicklung einer Token-Identitätsthese. Denn die Type/Token-Unterscheidung lässt sich plausiblerweise auch auf Eigenschaften anwenden. So lassen sich Eigenschaften als Typen von ihren konkreten Instantiierungen als Token unterscheiden.

[156] Neben Kims und Davidsons Ereigniskonzeptionen werden gelegentlich auch andere Ereigniskonzeptionen in Diskussionen des nicht-reduktiven Physikalismus vorausgesetzt. Eine Möglichkeit besteht etwa darin, Ereignisse über ihre essentiellen Eigenschaften zu individuieren (vgl. Yablo (1987), Yablo (1992) und Harbecke (2013)). Solche Konzeptionen haben mit Davidsons Konzeption gemein, dass Eigenschaften von Ereignissen instantiiert werden und nicht wie bei Kim konstitutiv für Ereignisse sind. Sie unterscheiden sich von Davidsons Konzeption jedoch darin, dass mehrere Ereignisse in derselben Raumzeitregion stattfinden können. Denn mehrere Ereignisse, die in derselben Raumzeitregion stattfinden und sogar dieselben aktualen Eigenschaften instantiieren, können sich dann alleine dadurch unterscheiden, dass ihnen unterschiedliche Eigenschaften essentiell zukommen. Sie unterscheiden sich dann allein in ihrem modalen Profil. So können mentale Ereignisse dadurch ausgezeichnet sein, dass ihnen mentale Eigenschaften essentiell zukommen und physische Ereignisse können dadurch ausgezeichnet sein, dass ihnen physische Eigenschaften essentiell zukommen. Mentale Ereignisse sind dann koinzident mit physischen Ereignissen, jedoch nicht identisch mit physischen Ereignissen: Denn sie können in Welten fortbestehen, in denen ihre essentielle mentale Eigenschaft auf physischer Ebene anders realisiert ist. Insofern den physischen Ereignissen der aktuale Realisierer essentiell zukommt, besteht das entsprechende physische Ereignis in den entsprechenden Welten nicht fort. Auch hier haben wir also eine Ablehnung der Ereignis-Identitätsthese, die mit dem nicht-reduktiven Physikalismus kompatibel ist.

[157] Vgl. z. B. Fodor (1974).

Es gibt beispielsweise die Eigenschaft, rot zu sein, und verschiedene Instantiierungen dieser Eigenschaft an einem roten Ball und einem roten Hemd usw. Vor diesem Hintergrund vertreten einige Autor*innen eine These der Nicht-Identität in Bezug auf Eigenschaften in Kombination mit einer Identitätsthese in Bezug auf konkrete Eigenschaftsinstantiierungen: Konkrete mentale und spezialwissenschaftliche Eigenschaftsinstantiierungen sind mit physischen Eigenschaftsinstantiierungen identisch, ohne dass die zugehörigen mentalen und spezialwissenschaftlichen Eigenschaften mit physischen Eigenschaften identisch sind.[158]

Eine solche Position kann innerhalb des nicht-reduktiven Physikalismus nach meiner Einschätzung durchaus konsistent vertreten werden. Prima facie spricht nichts dagegen, dass diese Position mit einem Realismus in Bezug auf mentale/spezialwissenschaftliche Eigenschaften und einer These der ontologischen Abhängigkeit vereinbart werden kann. Es hängt in ihrer Entwicklung jedoch einiges davon ab, wie das Verhältnis zwischen Eigenschaften und Eigenschaftsinstantiierungen bestimmt wird. Insbesondere muss ein und dieselbe Eigenschaftsinstantiierung eine Instantiierung mehrerer Eigenschaften sein können.[159] Es ist zumindest nicht unmittelbar einsichtig, wie dies sein kann. Denn naheliegend scheint es zunächst, Eigenschaftsinstantiierungen unter Bezug auf die Eigenschaften, deren Instantiierung sie sind, zu individuieren. Pro Eigenschaft, die zu einem Zeitpunkt an einem Objekt instantiiert ist, gibt es dann eine Eigenschaftsinstantiierung an diesem Objekt.[160]

Auch wenn also eine solche Token-Identitätstheorie innerhalb der Grenzen des nicht-reduktiven Physikalismus entwickelt werden kann, ist es keinesfalls plausibel, dass der nicht-reduktive Physikalismus auf eine solche Token-Identitätstheorie festgelegt ist. Denn wenn mentale/spezialwissenschaftliche Eigenschaften von physischen Eigenschaften in einem hinreichend starken Sinne ontologisch abhängen, besteht auch ein entsprechender enger Zusammenhang zwischen den einzelnen Instantiierungen mentaler/spezialwissenschaftlicher Eigenschaften und den Instantiierungen ihrer physischen Basen. Und dieser Zusammenhang dürfte bereits für die physikalistischen Verpflichtungen des nicht-reduktiven Physikalismus in Bezug auf Eigenschaftsinstantiierungen ausreichen.

---

[158] Vgl. z. B. Robb (1997).

[159] Macdonald und Macdonald (2006) verteidigen diese These.

[160] Für eine Kritik an der so verstandenen Token-Identitätstheorie, die detailliert auf die Frage des Verhältnisses zwischen Eigenschaften und Eigenschaftsinstantiierungen vor dem Hintergrund verschiedener Theorien von Eigenschaften eingeht, vgl. Schneider (2012). Van Riel (2014b) verteidigt die Token-Identitätstheorie gegen Schneiders Einwände.

Eine Identitätsthese in Bezug auf Eigenschaftsinstantiierungen muss also nicht vertreten werden.[161]

2.6.3 Anomaler Monismus

Auch Donald Davidsons anomaler Monismus[162] beruft sich an zentraler Stelle auf eine Ereignis-Identitätsthese. Dies macht den ‚monistischen' Teil des anomalen Monismus aus: Alle Ereignisse sind physisch. Wie schon in Abschnitt 2.6.1. diskutiert, ist diese These in Davidsons Verständnis jedoch zu schwach, um eine substantielle Verpflichtung des nicht-reduktiven Physikalismus auszudrücken. Dass alle Ereignisse – inklusive jener Ereignisse, die mentale/spezialwissenschaftliche Eigenschaften haben – auch physische Eigenschaften haben, ist damit kompatibel, dass mentale/spezialwissenschaftliche Eigenschaften beliebig unabhängig von physischen Eigenschaften sind. Es lässt sich jedoch noch mehr über das Verhältnis zwischen dem anomalen Monismus und dem nicht-reduktiven Physikalismus sagen. Aus verschiedenen Gründen stellt sich der anomale Monismus als eigenständige Position heraus, die nicht mit dem nicht-reduktiven Physikalismus verwechselt werden sollte.

Erstens beruht Davidsons Position auf einem recht radikalen Nominalismus: (Spärliche) Eigenschaften spielen in Davidsons Ontologie schlicht keine Rolle – ganz gleich, ob sie in einer nominalistisch akzeptablen Weise durch z. B. objektive Ähnlichkeitsrelationen oder in einer universalienrealistischen Weise aufgefasst werden. Eigenschaften gehören nicht zum Inventar von Davidsons Ontologie. Stattdessen gibt es nur Ereignisse und Objekte, die auf vielfältige Weisen sprachlich beschrieben werden können.[163] Wir haben ein Rohmaterial aus partikularen Objekten und Ereignissen und zahlreiche Möglichkeiten zur Beschreibung dieses Rohmaterials. Wir haben aber keine ‚vermittelnde' Entität in der Form von Universalien oder objektiven Ähnlichkeiten, die erklären

[161] Üblicherweise argumentieren die Vertreter*innen von solchen Token-Identitätstheorien, dass die Token-Identitätsthese für eine Erwiderung auf den Vorwurf des Epiphänomenalismus notwendig sei. Auch diese These lehne ich aber ab: In Teil 4 argumentiere ich dafür, dass eine andere Lösung für den Vorwurf des Epiphänomenalismus verfügbar ist.

[162] Vgl. Davidson (1970) und Davidson (1993).

[163] Vgl. z. B. Maurin (1998) für Diskussion.

würde, warum bestimmte Beschreibungen auf bestimmte partikulare Ereignisse oder Objekte zutreffen.[164]

Dieser nominalistische Hintergrund hat unter anderem die Konsequenz, dass die Klassifizierung von Ereignissen als physische Ereignisse oder mentale Ereignisse keine große ontologische Tragweite hat. Oben habe ich gesagt, dass ein Ereignis laut Davidson physisch ist, wenn es eine physische Eigenschaft hat und mental ist, wenn es eine mentale Eigenschaft hat. Zieht man den Nominalismus hinzu, laufen diese Bestimmungen darauf hinaus, dass ein Ereignis physisch ist, wenn man es physisch beschreiben kann und mental ist, wenn man es mental beschreiben kann. Dass man Ereignisse so beschreiben kann, hat nach Davidson jedoch nichts mit irgendwelchen substantiellen metaphysischen Merkmalen der Ereignisse zu tun. Dass alle Ereignisse physisch sind, ist dann mehr eine These über unsere Repräsentation der Welt als eine These über die Natur der Welt. Davidson könnte daher wohl treffender als neutraler Monist denn als Physikalist charakterisiert werden.[165]

Wenn man diesen radikalen Nominalismus für einen Moment ignoriert und Davidson so deutet, als rede er auch über Eigenschaften, denen irgendetwas in der Realität entspricht – seien es Universalien oder objektive Ähnlichkeiten – dann ergibt sich eine Position, die aus anderen Gründen nicht zum nicht-reduktiven Physikalismus passt: Denn eine metaphysisch notwendige Abhängigkeit mentaler Eigenschaften von physischen Eigenschaften verbietet sich bei Davidson aufgrund seiner Ablehnung physiopsychischer Gesetze. Diese macht den ‚anomalen' Teil des anomalen Monismus aus. Ein ontologisch (eigenschaftsrealistisch) gewendeter Davidson würde also den hier skizzierten nicht-reduktiven Physikalismus zugunsten einer stärker dualistischen oder emergentistischen Position ablehnen.[166]

---

[164] Vgl. Campbell (2008: 43): „Properties are, given Davidson's nominalism, purely a matter of description; […] Properly speaking, they are features of our theories about events rather than features of events themselves". Auch Heil (2009: 92) betont den nominalistischen Hintergrund des anomalen Monismus.

[165] In dieser Interpretation folge ich Heil (2009: 97).

[166] Dies mag in Konflikt damit stehen, dass Davidson (1970: 141) selbst eine Supervenienz-These in Betracht zieht. Tatsächlich scheint Davidsons Anomalie-These mit der Supervenienzthese in Konflikt zu stehen. Hier ist also eine interpretatorische Entscheidung zu treffen. Die relative Zentralität der Anomalie-These und die Beiläufigkeit der Supervenienz-These sprechen aber für die hier bevorzugte Auflösung.

Vgl. aber auch Davidson (1993: 7) und Child (1992) für Vereinbarungsversuche und Blackburn (1985) sowie Kim (1993a) für Betonungen des Konflikts.

2.6.4 Mentaler Exzeptionalismus: Haben mentale Eigenschaften einen Sonderstatus?

Als ‚mentalen Exzeptionalismus' möchte ich hier die These bezeichnen, dass mentale Eigenschaften mit Blick auf ihre ontologische Beziehung zu physischen Eigenschaften einen Sonderstatus im Vergleich mit anderen höherstufigen Eigenschaften haben. Laut dem mentalen Exzeptionalismus ist die Unterscheidung zwischen mentalen Eigenschaften auf der einen Seite und allen anderen höherstufigen Eigenschaften und den basalen physischen Eigenschaften auf der anderen Seite wichtiger und ontologisch ernster als die Unterscheidung zwischen allen höherstufigen (inklusive mentaler) Eigenschaften auf der einen Seite und den basalen physischen Eigenschaften auf der anderen Seite.

Es ist wichtig zu sehen, dass der nicht-reduktive Physikalismus *nicht* mit einem so verstandenen mentalen Exzeptionalismus einhergeht. Denn ein Sonderstatus mentaler Eigenschaften im nicht-reduktiven Physikalismus (im Vergleich zu spezialwissenschaftlichen Eigenschaften) ist eigentlich nicht zu haben. Die Argumente, die von nicht-reduktiven Physikalist*innen für eine Nicht-Identität mentaler Eigenschaften mit physischen Eigenschaften angeführt werden, sprechen auch für eine Nicht-Identität spezialwissenschaftlicher Eigenschaften mit physischen Eigenschaften (multiple Realisierung). Die Argumente, die für eine Auszeichnung mentaler Eigenschaften vor spezialwissenschaftlichen Eigenschaften sprechen (Qualia, Anomalie des Mentalen) sprechen zugleich gegen eine hinreichend enge Abhängigkeit mentaler Eigenschaften von physischen Eigenschaften und somit für einen genuinen Dualismus.

Innerhalb des Stufenmodells haben mentale Eigenschaften also keinen (oder zumindest nicht zwangsläufig einen) Sonderstatus. Ebenso wie zum Beispiel biologische und chemische Eigenschaften sind sie nicht mit physischen Eigenschaften identisch, weil sie auf physischer Ebene multipel realisiert sind. Die Unterscheidung zwischen fundamentalen, $\text{physischen}_{\text{eng}}$ Eigenschaften auf der einen Seite und höherstufigen Eigenschaften auf der anderen Seite ist also wichtiger als die Unterscheidung zwischen mentalen Eigenschaften auf der einen Seite und allen anderen Eigenschaften auf der anderen Seite.

2.6.5 Reduktiver und eliminativer Physikalismus

Hinsichtlich des reduktiven Physikalismus gilt es zunächst, zwei Versionen zu unterscheiden, die auf unterschiedliche Weisen vom nicht-reduktiven Physikalismus abgegrenzt sind:

Erstens gibt es Identitätstheorien.[167] Diese gehen von der Existenz spärlicher mentaler Eigenschaften aus und identifizieren sie mit physischen Eigenschaften. Von diesen Theorien ist der nicht-reduktive Physikalismus durch die These der Nicht-Identität abgegrenzt. In Bezug auf die These des Realismus können sich nicht-reduktive Physikalist*innen und Identitätstheoretiker*innen hingegen einig sein: Es gibt spärliche mentale Eigenschaften, die als Wahrmacher alltagspsychologischer Aussagen dienen. Identitätstheoretiker sind zugleich auf die Ablehnung jeder Form von multipler Realisierung festgelegt: Mentale Eigenschaften müssen mit den physischen Eigenschaften, mit denen sie identifiziert werden, notwendig ko-instantiiert sein.

Zweitens gibt es reduktive Physikalist*innen, die multiple Realisierung anerkennen und somit in gewisser Weise auch Eigenschaftsidentitäten ablehnen.[168] Diesen Versionen des reduktiven Physikalismus liegt letztlich eine Ablehnung des ontologischen Realismus in Bezug auf mentale Eigenschaften zugrunde: Es gibt keine (spärlichen) mentalen Eigenschaften, die als Wahrmacher alltagspsychologischer Aussagen fungieren. Stattdessen kommen wahre alltagspsychologische Aussagen ohne mentale Wahrmacher aus. Sie werden stattdessen von physischen Eigenschaften wahrgemacht. Diese Versionen des reduktiven Physikalismus sind also zugleich auf eine Ablehnung des starken semantischen Realismus festgelegt. Die Festlegungen auf eine Ablehnung des ontologischen Realismus und des starken semantischen Realismus unterscheiden sie vom nicht-reduktiven Physikalismus.

Eliminative Physikalist*innen[169] schließlich sehen keinen Bedarf für Wahrmacher alltagspsychologischer Aussagen, weil sie diese Aussagen für falsch halten. Sie können daher an einem starken semantischen Realismus festhalten. Sie sind durch die Ablehnung des ontologischen Realismus sowie die für sie spezifische These der Falschheit alltagspsychologischer Aussagen ausgezeichnet.

Diese knappe Diskussion zeigt bereits, dass der nicht-reduktive Physikalismus durch die spezifische Kombination der These des Realismus und der These der Nicht-Identität klar von anderen physikalistischen Positionen abgegrenzt ist.

---

[167] Vgl. klassisch Place (1956) und Smart (1959) sowie Polger (2009) und Gozzano und Hill (2012) für neuere Verteidigungen der Identitätstheorie.

[168] Vgl. Lewis (1994), Heil (2003), Ney (2016a) und Morris (2019). Auch Kims (2005) Position fällt nach meiner Einschätzung in diese Klasse von Theorien: Funktionale Reduktionen multipel realisierter Eigenschaften eliminieren multipel realisierte Eigenschaften, während sie an der Wahrheit der zugehörigen Aussagen festhalten.

[169] Vgl. Churchland (1981) und Churchland (1993) für Verteidigungen sowie Baker (1988) und Hannan (1993) für Kritik.

Der nicht-reduktive Physikalismus kollabiert nicht in eine andere Version des Physikalismus.[170]

2.6.6 Naturalistischer Dualismus

Dualistische Theorien mentaler Eigenschaften teilen mit dem nicht-reduktiven Physikalismus die folgenden nicht-reduktiven Kernthesen in Bezug auf mentale Eigenschaften: Mentale Eigenschaften sind real und nicht mit physischen Eigenschaften identisch.[171] Jedoch lehnen sie die These der ontologischen Abhängigkeit ab: Laut dem Dualismus sind mentale Eigenschaften nicht alleine *aufgrund von* physischen Eigenschaften instantiiert. Sie sind ‚mehr als' physische Eigenschaften. Und in der dualistischen Version des fiktiven Schöpfungsmythos musste Gott neben der Verteilung der physischen Eigenschaften auch noch die mentalen Eigenschaften verteilen. Mentale Eigenschaften sind laut dem Dualismus nicht in einer physikalistisch akzeptablen Weise von physischen Eigenschaften ontologisch abhängig.

Zugleich ist der Dualismus nicht darauf festgelegt, dass mentale Eigenschaften vollkommen losgelöst von physischen Eigenschaften sind. Vielmehr sieht die vielleicht plausibelste Version des Dualismus – der *naturalistische Dualismus*[172] – durchaus einen engen Zusammenhang zwischen mentalen und physischen Eigenschaften. Nur ist der behauptete Zusammenhang eben nicht *so eng* wie im nicht-reduktiven Physikalismus. Wie lässt sich diese Trennung präzisieren?

Üblich und plausibel ist eine Abgrenzung über die modale Stärke der Abhängigkeitsrelation:[173] Wie in 2.5.7. geschildert, ist der nicht-reduktive Physikalismus auf eine These der universalen metaphysischen Necessitation festgelegt: Jede instantiierte mentale Eigenschaft wird durch eine instantiierte physische Eigenschaft *mit metaphysischer Notwendigkeit* necessitiert. In allen metaphysisch möglichen Welten, in denen eine totale physische Basis der mentalen Eigenschaft **M** instantiiert ist, ist auch **M** instantiiert. Naturalistische Dualisten können eine modal schwächere Necessitationsthese akzeptieren: Jede instantiierte mentale

[170] Auch Haug (2011) bezieht sich in seiner Diskussion der Unterscheidung zwischen reduktivem und nicht-reduktivem Physikalismus in erster Linie auf realistische Thesen.

[171] Naheliegend wäre auch eine Akzeptanz der These der multiplen Realisierung. Dies ist jedoch sicherlich keine notwendige Bedingung für den Dualismus. Die wichtigsten Argumente, die für den Dualismus sprechen, kommen ohne einen Bezug auf multiple Realisierung aus.

[172] Vgl. Chalmers (2017) und Kroedel (2019: 18).

[173] Vgl. z. B. Kroedel (2019: 18).

Eigenschaft wird durch eine instantiierte physische Eigenschaft *mit nomologischer Notwendigkeit* necessitiert. In allen Welten, in denen eine totale physische Basis_{D}[174] von **M** instantiiert ist *und in der dieselben Naturgesetze gelten wie in der aktualen Welt*, ist **M** instantiiert.

In 2.5.7 habe ich auch darauf hingewiesen, dass eine totale physische Basis für eine mentale Eigenschaft im nicht-reduktiven Physikalismus recht komplex sein kann und insbesondere einen Verweis auf physische Naturgesetze beinhalten könnte. Damit es sich um eine totale *physische* Basis handelt, sollten solche Verweise aber auch auf *physische* Naturgesetze beschränkt bleiben. Eine totale physische Basis für Schmerz könnte – wenn ein Verweis auf physische Gesetze benötigt wird[175] – eine komplexe Eigenschaft der folgenden Art sein: ‚ein C-Faser-Feuern[176] in der richtigen physischen Umgebung unter Gelten der relevanten physischen Naturgesetze sein'. Nicht-reduktive Physikalist*innen können dann behaupten, dass diese totale physische Basis Schmerz metaphysisch necessitiert. Naturalistische Dualisten würden diese Behauptung ablehnen.

Denn – so die Idee – die *physischen* Naturgesetze, auf die in der Spezifikation der totalen Basis verwiesen wird, sind für die Instantiierung der *mentalen* Eigenschaft nicht direkt relevant. Die mentale Eigenschaft ergibt sich erst aufgrund eines zusätzlichen, nicht rein-physischen Naturgesetzes. Ein solches physisch-mentales Gesetz wird die Instantiierung einer physischen Eigenschaft gesetzesmäßig mit der Instantiierung der zugehörigen mentalen Eigenschaft verknüpfen. Die Idee ist hier also, dass es neben physischen Gesetzen auch noch *grundlegende physisch-mentale Gesetze* gibt. In metaphysisch möglichen Welten, in denen diese physisch-mentalen Gesetze nicht gelten, kann die totale Basis_{D} einer mentalen Eigenschaft ohne die zugehörige mentale Eigenschaft instantiiert sein. Die totale Basis_{D} geht also nur in allen Welten, in denen dieselben Naturgesetze (inklusive der physisch-mentalen Gesetze) gelten wie in der aktualen Welt, mit der zugehörigen mentalen Eigenschaft einher. Deshalb ist die

[174] Der Index soll darauf hinweisen, dass hier ein anderer Begriff einer totalen Basis im Spiel ist, als der in 2.5.7. eingeführte. Dort habe ich den Begriff einer totalen Basis an metaphysische Necessitation gebunden. Für Dualisten müsste der Begriff an nomologische Necessitation gebunden werden.

[175] Es könnte eben auch reichen, dass die physischen Eigenschaften in einer geeigneten Abfolge instantiiert sind, ohne dass sie dies *gesetzmäßig* tun. Insbesondere unter Voraussetzung einer Hume'schen Konzeption von Naturgesetzen und einer ausreichend weiträumigen Bestimmung der physischen Eigenschaftsinstantiierungen könnte ein zusätzlicher Verweis auf Gesetze zudem redundant sein.

[176] Um eine $\text{physische}_{\text{eng}}$ Eigenschaften zu gewinnen müsste man hier strenggenommen eine komplexere Eigenschaft einsetzten. Für den Punkt, den ich mit dem Beispiel illustrieren will, ist das jedoch nicht relevant.

Necessitation im naturalistischen Dualismus nur mit nomologischer und nicht mit metaphysischer Notwendigkeit verbunden.[177]

Der Abgrenzung des nicht-reduktiven Physikalismus vom naturalistischen Dualismus über die modale Stärke der Necessitationsbehauptung liegt also eine These über die Existenz verschiedener Arten von grundlegenden Naturgesetzen zugrunde. Während nicht-reduktive Physikalist*innen meinen, dass alle grundlegenden Naturgesetze[178] physisch sind, glauben naturalistische Dualisten an zusätzliche grundlegende physisch-mentale Naturgesetze.

Gegen die Abgrenzung des nicht-reduktiven Physikalismus vom naturalistischen Dualismus über die modale Stärke der Necessitationsbehauptung gibt es zwei Einwände, die ich im Folgenden noch kurz diskutieren möchte:

Erstens bricht die Unterscheidung über die modale Stärke der Necessitationsbehauptung zusammen, wenn die aktualen Naturgesetze metaphysisch notwendig sind.[179] Unter dieser Voraussetzung fällt die Menge der metaphysisch möglichen Welten mit der Menge der nomologisch möglichen Welten zusammen. Sowohl naturalistische Dualist*innen als auch nicht-reduktive Physikalist*innen würden dann eine metaphysische Necessitationsthese vertreten.

In Erwiderung auf diesen Einwand kann man darauf verweisen, dass es – selbst unter der Voraussetzung, dass die aktualen Naturgesetze metaphysisch notwendig sind – noch immer einen Unterschied zwischen naturalistischem Dualismus und nicht-reduktivem Physikalismus mit Blick auf die Frage gibt, was die aktualen Naturgesetze *sind*. Naturalistische Dualist*innen behaupten die Existenz grundlegender physisch-mentaler Naturgesetze, während nicht-reduktive Physikalist*innen diese Behauptung ablehnen. Wenn die aktualen Naturgesetze metaphysisch notwendig sind, muss der nicht-reduktive Physikalismus also durch die für den naturalistischen Dualismus spezifische These der Existenz grundlegender physisch-mentaler Naturgesetze abgegrenzt werden.

---

[177] Im naturalistischen Dualismus gibt es natürlich metaphysisch necessitierende totale Basen, in denen ein Verweis auf die relevanten physischen *und physisch-mentalen* Naturgesetze kodiert ist (‚ein C-Faser Feuern in der geeigneten Umgebung unter Gelten der relevanten physischen und physisch-mentalen Naturgesetze sein). Dies würde jedoch nicht mehr als totale *physische* Basis zählen.

[178] Hier ist Vorsicht geboten: Nicht-reduktive Physikalist*innen meinen durchaus, dass es höherstufige Gesetze gibt. Auch diese ergeben sich jedoch mit Notwendigkeit aus der physischen Beschaffenheit der Welt und den physischen Gesetzen. In diesem Sinne sind die höherstufigen Gesetze nicht grundlegend. Vgl. für Diskussion dieses Punkts z. B. Loewer (2009).

[179] Vgl. Wilson (2005: 436–454).

Zweitens ist die These der metaphysischen Necessitation für sich genommen wohl nicht stark genug, um die für den nicht-reduktiven Physikalismus ausschlaggebenden ‚aufgrund von' und ‚nicht mehr als'-Thesen zu begründen. Dies eröffnet die prinzipielle Möglichkeit einer Position, die die metaphysische Necessitationsthese akzeptiert und die entsprechenden ‚aufgrund von'- und ‚nicht mehr als'-Thesen ablehnt. Obgleich mentale Eigenschaften dann von physischen Eigenschaften metaphysisch necessitiert werden, sind sie dennoch mehr als physische Eigenschaften und nicht aufgrund von physischen Eigenschaften instantiiert. Eine solche (gedachte) Position kann als ‚Necessitations-Dualismus' bezeichnet werden. Das Problem besteht dann darin, dass sich der nicht-reduktive Physikalismus nicht vom Necessitations-Dualismus abgrenzen kann.[180]

Zwei Erwiderungen sind hier aussichtsreich:
Erstens ist es keine ausgemachte Sache, dass der Necessitationsdualismus eine konsistente Position ist. Die Betonung des ‚Neuen' mentaler Eigenschaften müsste erst einmal auf überzeugende Weise mit der Necessitationsthese in Einklang gebracht werden. Die bekannten Gegenbeispiele gegen die Verbindung von Necessitation und ‚aufgrund von'- oder ‚nicht mehr als'-Behauptungen betreffen z. B. Selbstnecessitation (A necessitiert A, aber ist nicht aufgrund von A instantiiert) oder notwendig instantiierte Eigenschaften (Die Eigenschaft, rot-oder-nicht-rot zu sein, wird trivialerweise von jeder beliebigen Eigenschaft necessitiert, ist aber nicht trivialerweise aufgrund von jeder beliebigen Eigenschaft instantiiert). Diese beiden Sorten von Gegenbeispielen funktionieren mit Bezug auf mentale Eigenschaften sicherlich nicht. Mentale Eigenschaften können im Rahmen eines Necessitations-Dualismus nicht mit physischen Eigenschaften identisch sein. Ein Fall von Selbstnecessitation liegt also nicht vor. Auch sind mentale Eigenschaften nicht notwendig instantiiert. Auch dieser Grund steht also nicht zur Verfügung, um die These zu untermauern, dass mentale Eigenschaften von physischen Eigenschaften necessitiert werden und dennoch *nicht* aufgrund von physischen Eigenschaften instantiiert werden.[181]

[180] Vgl. Stoljar (2010: 144–150).

[181] Man könnte mir hier vielleicht unfaire Beweislastverschiebung vorwerfen. Es ist jedoch durchaus plausibel, dass sich der nicht-reduktive Physikalismus nur von ernstzunehmenden Positionen abzugrenzen hat. Insofern bloß auf die prinzipielle Vorstellbarkeit einer eventuell zu entwickelnden Position verwiesen wird, die sich auf einem freien Platz im logischen Raum ansiedeln ließe, halte ich den Einwand aus dem Necessitations-Dualismus nicht für sehr stark.

Sollte dieser Erwiderung nicht zufriedenstellen, können nicht-reduktive Physikalist*innen zwecks Abgrenzung vom Necessitations-Dualismus zweitens zu einer der substantielleren Relationen ontologischer Abhängigkeit aus Abschnitt 2.5. Zuflucht nehmen. Insofern diese Relationen – Realisierung, Determination, oder Grounding – die gesuchten ‚aufgrund von' und ‚nicht mehr als'-Behauptungen begründen können, werden sie von Necessitations-Dualist*innen in jedem Falle abgelehnt.

Insoweit die Argumentation dieser Arbeit betroffen ist, gilt es noch folgende Anmerkung zu machen: Wenn es tatsächlich dualistische Positionen gibt, die die metaphysische Necessitationsthese akzeptieren können, dann sitzen diese Versionen des Dualismus mit Blick auf den Vorwurf des Epiphänomenalismus mit dem nicht-reduktiven Physikalismus im selben Boot. Denn Schlüssel zu einer überzeugenden Erwiderung auf den Vorwurf des Epiphänomenalismus ist nach der vorliegenden Argumentation die Necessitationsthese und nicht eine darüber hinausgehende erklärende Beziehung ontologischer Abhängigkeit. Der Grundgedanke dabei ist, dass es in kausalen Kontexten in erster Linie auf modale Verknüpfungen ankommt und nicht so sehr auf die Erklärung der modalen Verknüpfung.

2.6.7 Nicht-reduktiver Physikalismus und Positionen zum Status mentaler und physischer Verursachung

Ich habe den nicht-reduktiven Physikalismus in diesem Teil gänzlich ohne Bezug auf kausale Thesen bestimmt. Dies ist eher ungewöhnlich. In der Literatur finden sich häufig Definitionen des nicht-reduktiven Physikalismus, die eine oder beide der folgenden Thesen als notwendige Bedingungen für den nicht-reduktiven Physikalismus anführen:[182]

Mentale/spezialwissenschaftliche Verursachung: Mentale/spezialwissenschaftliche Ereignisse sind Ursachen für physische und andere mentale/spezialwissenschaftliche Ereignisse.

Kausale Geschlossenheit: Physische Ereignisse haben hinreichende physische Ursachen (insofern sie überhaupt hinreichende Ursachen haben).

[182] Als willkürlich ausgewählte Beispiele seien hier Kim (2005: 35), Baker (2009: 112) und List und Menzies (2009: 475) genannt.

Beide Thesen werden in Teil 3 dieser Arbeit natürlich noch ausführlich behandelt. Ich möchte an dieser Stelle jedoch kurz erklären, warum ich sie nicht als notwendige Bedingungen für den nicht-reduktiven Physikalismus auffasse. Ich habe hierfür im Wesentlichen drei Gründe:

Erstens: Ich sehe den nicht-reduktiven Physikalismus in erster Linie als eine Position zum ontologischen Status mentaler/spezialwissenschaftlicher Eigenschaften und ihrem (nicht-kausalen) Verhältnis zu physischen Eigenschaften. Als solcher muss er sich von Alternativen wie dem Dualismus und dem reduktiven oder eliminativen Physikalismus abgrenzen. Wie in den letzten Abschnitten deutlich wurde, ist diese Abgrenzung möglich, ohne auf die genannten kausalen Thesen zu verweisen. Der Kern des nicht-reduktiven Physikalismus wird daher besser ohne Verweis auf die kausalen Thesen eingefangen.

Zweitens: Ich sehe keinen Grund, warum der nicht-reduktive Physikalismus mit einem Skeptizismus mit Blick auf Kausalität in der Physik oder gar mit einem allgemeinen Kausalitätsskeptizismus inkompatibel sein sollte. Die Überlegungen, die für eine dieser Positionen sprechen, haben mit dem Verhältnis zwischen mentalen/spezialwissenschaftlichen Eigenschaften und physischen Eigenschaften nicht viel zu tun. Wer sich beispielsweise von Ernst Machs oder Betrand Russells Argumenten für die Elimination kausalen Vokabulars aus der Physik überzeugen lässt, wird die These der kausalen Geschlossenheit ablehnen.[183] Wer Kausalität begrifflich an notwendige Verknüpfung zwischen distinkten Ereignissen bindet und solche notwendigen Verknüpfungen als gute Humeaner*in ablehnt, wird auch die These der mentalen/spezialwissenschaftlichen Verursachung ablehnen. Es ist jedoch nicht zu sehen, weshalb solche Überlegungen zu einer Ablehnung des nicht-reduktiven Physikalismus als einer Position über den Status mentaler/spezialwissenschaftlicher Eigenschaften in einer physischen Welt führen sollten.

Drittens: Die Trennung des Kerns des nicht-reduktiven Physikalismus von den genannten kausalen Thesen erlaubt eine bessere Aufteilung des logischen Raumes von Positionen. Denn die Frage nach dem ontologischen Status mentaler und spezialwissenschaftlicher Eigenschaften ist – zumindest auf den ersten Blick – logisch unabhängig von der Frage, in welche Kausalbeziehungen mentale und spezialwissenschaftliche Eigenschaften eintreten. Eine Unterscheidung zwischen epiphänomenalistischen, parallelistischen und interaktionistischen Positionen, die sich an der Frage nach der kausalen Rolle mentaler Eigenschaften orientiert, kann deshalb sowohl innerhalb von dualistischen als auch innerhalb von (reduktiven

[183] Vgl. Mach (1896), Russell (1912), Price und Corry (2007) und Hüttemann (2013a: 48–53).

oder nicht-reduktiven) physikalistischen Positionen getroffen werden.[184] Ein epiphänomenalistischer nicht-reduktiver Physikalismus ist ebenso möglich wie ein interaktionistischer oder ein parallelistischer nicht-reduktiver Physikalismus. Es gibt keinen Grund, den epiphänomenalistischen nicht-reduktiven Physikalismus schon durch die Definition des Ausdrucks ‚nicht-reduktiver Physikalismus' auszuschließen. Und es wäre begrifflich umständlicher, wenn man eine Position, die die hier vorausgesetzten Kernthesen vertritt aber die Existenz mentaler Verursachung ablehnt, neu benennen müsste, weil der Begriff ‚nicht-reduktiver Physikalismus' definitionsbedingt nicht auf sie zutrifft.



[184] Auch ein epiphänomenalistischer reduktiver Physikalismus ist zumindest konsistent. Es spricht begrifflich nichts dagegen, mentale Eigenschaften mit epiphänomenalen physischen Eigenschaften zu identifizieren. Dies könnte unter Umständen bei Positionen herauskommen, denen zufolge mentale Eigenschaften mit disjunktiven physischen Eigenschaften identisch sind.



3 Nicht-reduktiver Physikalismus und mentale Verursachung

Der nicht-reduktive Physikalismus wirkt deshalb attraktiv, weil er einerseits realistischen Intuitionen in Bezug auf mentale Entitäten gerecht wird, andererseits mentale Entitäten aber auch in einer grundlegend physischen Welt verortet. Auf diese Weise verspricht er eine Vermittlung zwischen unseren alltäglichen, vortheoretischen Überzeugungen über denkende und handelnde Personen auf der einen Seite und dem theoretischen Bedürfnis nach einer einfachen, einheitlichen und wissenschaftsnahen Metaphysik auf der anderen Seite.

In der bisherigen Charakterisierung des nicht-reduktiven Physikalismus habe ich jedoch einen wichtigen Aspekt ausgeklammert: Zu unseren vor-theoretischen Überzeugungen gehört es nach allgemeiner Auffassung nicht nur, dass mentale Entitäten einfach bloß existieren. Vielmehr müssen sie auch wirksam sein. Sie müssen den Verlauf der Welt mitbestimmen und eine Funktion erfüllen. Personen müssen nicht nur faktisch denken und fühlen. Ihre Gedanken und Gefühle müssen auch einen Unterschied machen. Kurz: Es muss mentale Verursachung geben.

An dieser Stelle setzt nun der stärkste und meistdiskutierte Einwand gegen den nicht-reduktiven Physikalismus an: der Vorwurf des Epiphänomenalismus. Laut diesem Vorwurf kann der nicht-reduktive Physikalismus der Existenz mentaler Verursachung nicht Rechnung tragen. Da es sich bei der Annahme mentaler Verursachung zugleich um eine absolut zentrale und unverzichtbare Annahme handele, sei der nicht-reduktive Physikalismus aufzugeben. Er führe unweigerlich in die absurde Position des Epiphänomenalismus, demzufolge es keine mentale Verursachung gibt.

Ziel des zweiten Teils dieser Arbeit ist es, den Vorwurf des Epiphänomenalismus ausführlich zu rekonstruieren und zu diskutieren. Dabei gehe ich wie folgt vor: In Abschnitt 3.1. gehe ich zunächst knapp auf die These der mentalen Verursachung ein und skizziere den Vorwurf des Epiphänomenalismus, demzufolge nicht-reduktive Physikalist*innen die These der mentalen Verursachung

M. Rolffs, *Kausalität und mentale Verursachung*, Abhandlungen zur Philosophie,
https://doi.org/10.1007/978-3-662-66778-1_3

nicht akzeptieren können, obgleich jeder diese These akzeptieren muss. Die restlichen Abschnitte konstituieren dann eine ausführliche Diskussion des Vorwurfs des Epiphänomenalismus. In Abschnitt 3.2. gehe ich ausführlich auf die These der kausalen Geschlossenheit ein, die eine tragende Rolle im Vorwurf des Epiphänomenalismus spielt. Darauf aufbauend diskutiere ich in den Abschnitten 3.3. und 3.4. Argumente dafür, dass nicht-reduktive Physikalist*innen die These der mentalen Verursachung nicht akzeptieren können. In Abschnitt 3.5. schließlich diskutiere ich Argumente dafür, dass jeder die These der mentalen Verursachung akzeptieren muss.

3.1 Mentale Verursachung und der Vorwurf des Epiphänomenalismus

In diesem Abschnitt führe ich den Vorwurf des Epiphänomenalismus ein. In Abschnitt 3.1.1. führe ich anhand eines Beispiels vier Arten von mentaler Verursachung ein. In Abschnitt 3.1.2. unterscheide ich epiphänomenalistische und kausalistische Versionen des nicht-reduktiven Physikalismus. Abschnitt 3.1.3. widmet sich einer knappen Rekonstruktion des Vorwurfs des Epiphänomenalismus.

3.1.1 Mentale Verursachung: Ein Beispiel und vier Thesen

Um ein Gefühl dafür zu bekommen, welche Phänomene mit dem Begriff ‚mentale Verursachung' angesprochen sind, kann man sich das folgende, einfache Beispiel vor Augen halten: Susi sitzt in einem Philosophieseminar, hat einen interessanten Gedanken, und möchte diesen in einem Wortbeitrag mitteilen. Sie hebt daher ihre Hand, um die Aufmerksamkeit der Dozentin auf sich zu lenken. In dieses Beispiel lassen sich nun eine Reihe von Thesen der mentalen Verursachung hineinlesen.

> **Mental-mentale Verursachung**: Susis interessanter Gedanke verursacht ihren Wunsch, die Aufmerksamkeit der Dozentin auf sich zu lenken. Hier haben wir einen Fall *mental-mentaler Verursachung*: Ein mentales Ereignis (Susis Gedanke) verursacht ein anderes mentales Ereignis (Susis Wunsch).
>
> **Mental-behaviorale Verursachung**: Susis Wunsch, die Aufmerksamkeit der Dozentin auf sich zu lenken, verursacht, dass Susi ihren Arm hebt. Hier verursacht ein mentales Ereignis (Susis Wunsch) eine Körperbewegung oder eine Verhaltensweise (ihr Armheben). Ich spreche hier von *mental-behavioraler Verursachung*.

Mental-neurologische Verursachung: Susis Wunsch verursacht (womöglich über weitere neurologische Zwischenglieder) bestimmte neurologische Vorgänge in ihrem motorischen Cortex, die schließlich dazu führen, dass sie ihren Arm hebt. Laut dieser These verursacht ein mentales Ereignis (Susis Wunsch) neurologische Ereignisse (Vorgänge im motorischen Cortex). Ich spreche in so einem Fall allgemein von *mental-neurologischer Verursachung*.

Mental-mikrophysikalische Verursachung: Susis Wunsch verursacht Veränderungen in den Elementarteilchen, die die Neuronen in ihrem motorischen Cortex konstituieren. Laut dieser These verursacht ein mentales Ereignis (Susis Wunsch) mikrophysikalische Ereignisse. Hier spreche ich von *mental-mikrophysikalischer Verursachung*.

All dies sind Fälle von mentaler Verursachung. Im Allgemeinen spreche ich immer dann von mentaler Verursachung, wenn ein mentales Ereignis als Ursache in Erscheinung tritt.

Ich habe oben geschrieben, dass sich diese Behauptungen in das Beispiel ‚hineinlesen' lassen. Diese vage Formulierung lässt absichtlich offen, ob sich diese Behauptungen zwingend aus dem Beispiel ergeben. Folgt daraus, dass Susi ihre Hand hebt, um die Aufmerksamkeit der Dozentin auf sich zu lenken, tatsächlich, dass es mental-behaviorale, mental-neurologische und mental-mikrophysikalische Verursachung gibt? Diese Fragen nehme ich erst in Abschnitt 3.5 wieder auf. An dieser Stelle dient das Beispiel lediglich einer ersten Veranschaulichung der These der Existenz mentaler Verursachung.

3.1.2 Epiphänomenalismus und Kausalismus

Da der nicht-reduktive Physikalismus nach meiner Charakterisierung in Teil 2 dieser Arbeit nicht per definitionem auf die Existenz mentaler Verursachung festgelegt ist,[1] lassen sich verschiedene Versionen des nicht-reduktiven Physikalismus mit Blick auf ihre Einstellung zu mentaler Verursachung unterscheiden. Insbesondere ergibt sich zunächst eine einfache Unterscheidung zwischen einem epiphänomenalistischen nicht-reduktiven Physikalismus und einem kausalistischen nicht-reduktiven Physikalismus.

Der *epiphänomenalistische nicht-reduktive Physikalismus* (im Folgenden schlicht ‚*Epiphänomenalismus*') ergänzt die Kernthesen des nicht-reduktiven Physikalismus durch die Annahme, dass es keine mentale Verursachung gibt. Dem

[1] Wie bereits angemerkt weiche ich hiermit von anderen, verbreiteten Formulierungen des nicht-reduktiven Physikalismus ab.

Epiphänomenalismus zufolge sind also alle oben angeführten Behauptungen mentaler Verursachung falsch: Susis Gedanke verursacht nicht ihren Wunsch und ihr Wunsch verursacht weder ihr Verhalten, noch die Vorgänge in ihrem motorischen Cortex oder die Veränderungen in den Elementarteilchen, die ihren motorischen Cortex konstituieren.

Der *kausalistische nicht-reduktive Physikalismus* (im Folgenden schlicht ‚*Kausalismus*') ergänzt die Kernthesen des nicht-reduktiven Physikalismus durch die Annahme, dass es mentale Verursachung gibt. Der Kausalismus ist dabei als allgemeine Position neutral gegenüber der Frage, welche konkreten Behauptungen mentaler Verursachung wahr sind. Mit dem Kausalismus ist es zum Beispiel auch vereinbar, dass es nur mental-mentale Verursachung gibt, aber keine mental-behaviorale oder mental-neurologische Verursachung. Es ergibt sich also die Möglichkeit weiterer Unterscheidungen innerhalb des Kausalismus. Diese werde ich später (in Abschnitt 3.4.) auch wieder aufgreifen.

3.1.3 Der Vorwurf des Epiphänomenalismus

Der Vorwurf des Epiphänomenalismus ist der meistdiskutierte Einwand gegen den nicht-reduktiven Physikalismus. Laut diesem Vorwurf kann der nicht-reduktive Physikalismus der Existenz mentaler Verursachung nicht Rechnung tragen. Da es sich bei der Annahme mentaler Verursachung zugleich um eine absolut zentrale und unverzichtbare Annahme handele, sei der nicht-reduktive Physikalismus aufzugeben. Der nicht-reduktive Physikalismus führe unweigerlich in den Epiphänomenalismus, der aber keine haltbare Position sei.

Eine einfache Rekonstruktion des Vorwurfs des Epiphänomenalismus ist also die folgende:

Vorwurf des Epiphänomenalismus

(VE1) Wenn der nicht-reduktive Physikalismus wahr ist, dann gibt es keine mentale Verursachung.

(VE2) Es gibt mentale Verursachung.

(VEC) Der nicht-reduktive Physikalismus ist falsch.

Gegeben diese einfache Struktur des Vorwurfs des Epiphänomenalismus gibt es zwei grundsätzliche Strategien, den nicht-reduktiven Physikalismus zu verteidigen. Die *kausalistische Erwiderung* lehnt Prämisse (VE1) ab. Für Kausalist*innen

besteht die Herausforderung also darin, die Argumente, die zugunsten von (VE1) vorgebracht werden, zurückzuweisen. Das prominenteste Argument für Prämisse (VE1) ist dabei Jaegwon Kims Exklusionsargument. In Abschnitt 3.3. werde ich verschiedene Möglichkeiten erkunden, dieses Argument im Detail zu entwickeln. In Abschnitt 3.4. gehe ich dann auf kausalistische Erwiderungen auf dieses Argument ein.

Die *epiphänomenalistische Erwiderung* auf den Vorwurf des Epiphänomenalismus lehnt Prämisse (VE2) ab. Epiphänomenalist*innen stehen vor der Herausforderung, den Epiphänomenalismus zu plausibilisieren. Dies ist auf den ersten Blick keine einfache Aufgabe: Natürlich – so könnte man meinen – verursacht Susis Wunsch, die Aufmerksamkeit der Dozentin auf sich zu lenken, ihre Armbewegung. In Abschnitt 3.5. gehe ich auf die Aussichten ein, den Epiphänomenalismus trotz dieses Bedenkens zu verteidigen.

3.2 Die These der kausalen Geschlossenheit

In einer groben, ersten Charakterisierung besagt die These der kausalen Geschlossenheit, dass physische Ereignisse im Allgemeinen hinreichende physische Ursachen haben. Diese These nimmt auch im Vorwurf des Epiphänomenalismus eine zentrale Rolle ein. Sie führt zu einem wichtigen Argument für den ersten Schritt des Vorwurfs des Epiphänomenalismus. Bevor ich (in Abschnitt 3.3) auf dieses Argument eingehe und die Rolle der These der kausalen Geschlossenheit in diesem Argument diskutiere, ist es daher sinnvoll, etwas genauer auf die These der kausalen Geschlossenheit selbst einzugehen. Was besagt die These der kausalen Geschlossenheit? Welche impliziten Annahmen über die Natur der Kausalrelation enthält sie? Und sind nicht-reduktive Physikalist*innen tatsächlich zur Akzeptanz der These verpflichtet?

Um diese Fragen zu beantworten, gehe ich wie folgt vor: In Abschnitt 3.2.1. formuliere und präzisiere ich die These der kausalen Geschlossenheit. In Abschnitt 3.2.2. diskutiere ich einige Argumente zugunsten der These der kausalen Geschlossenheit. Aufbauend hierauf behandle ich in Abschnitt 3.2.3. die Frage, welche Reichweite die These der kausalen Geschlossenheit hat – auf welchen Bereich von Ereignissen sie sich also bezieht. In Abschnitt 3.2.4. arbeite ich einige implizite Voraussetzungen der These heraus, die die Natur der Verursachungsrelation betreffen. Abschnitt 3.2.5. schließlich behandelt die Rolle der These der kausalen Geschlossenheit innerhalb des nicht-reduktiven Physikalismus.

3.2.1 Die These der kausalen Geschlossenheit: Grundidee und Formulierungen

Ein Ausdruck der Grundidee der These der kausalen Geschlossenheit findet sich bei Kim:

> According to this principle, physics is causally and explanatorily self-sufficient: there is no need to go outside the physical domain to find a cause, or a causal explanation, of a physical event.[2]

Die These macht also eine Einschränkung geltend, die die kausale Geschichte physischer Ereignisse betrifft: Verfolgt man die Ursachenkette zurück, die zu einem beliebigen physischen Ereignis führt, muss man den Bereich des Physischen nicht verlassen. Man wird immer andere physische Ereignisse finden, die das physische Ereignis verursachen. Diese Einschränkung wird zugleich mit der These in Verbindung gebracht, dass die Physik kausal und explanatorisch eigenständig ist. Physikalische Erklärungen sind an keiner Stelle auf die Erwähnung nicht-physischer Ursachen angewiesen.

Diese Grundidee kann nun auf verschiedene Weisen präzisiert werden. In der Literatur findet sich eine große Vielfalt an nicht-äquivalenten Formulierungen der These der kausalen Geschlossenheit.[3] Im Folgenden gehe ich auf einige dieser Formulierungen ein.

Ein erster Versuch der Präzisierung könnte etwa in der folgenden These bestehen:

> **(KG1)** Alle physischen Ereignisse haben physische Ursachen.[4]

Diese These bringt jedoch nicht gut zum Ausdruck, was üblicherweise unter der These der kausalen Geschlossenheit verstanden wird. Dies liegt unter anderem daran, dass These (KG1) in einem Sinne sehr schwach ist: Sie legt fest, dass physische Ereignisse auch physische Ursachen haben. Sie erlaubt aber,

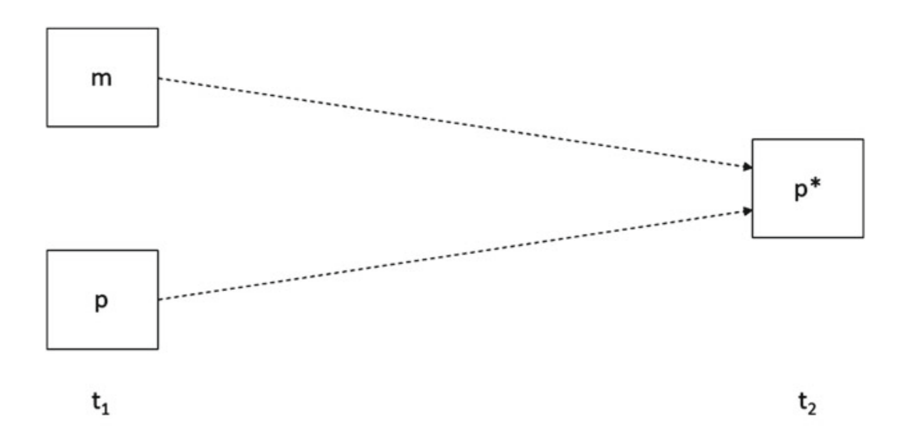
---

[2] Kim (2005: 16).

[3] Vgl. Gibb (2015: 628) für einige Beispiele unterschiedlicher Formulierungen. Vgl. auch Lowe (2000: 574): „One might have hoped for more exactitude and agreement amongst physicalists when it comes to the formulation of a principle so central to their position.“

[4] Diese genaue Formulierung ist aufgrund der im Folgenden auszuführenden Schwächen unüblich. Eine Formulierung, die dieser Formulierung zumindest in der Hinsicht gleicht, dass von ‚Ursachen‘ und nicht von ‚hinreichenden Ursachen‘ die Rede ist, findet sich jedoch z. B. in Kim (2005: 15): „If a a physical event has a cause at t, then it has a physical cause at t.“

dass diese physischen Ursachen für sich genommen nicht hinreichen, um ein physisches Ereignis hervorzubringen. Sie ist kompatibel damit, dass physische Ursachen nur im Verbund mit nicht-physischen Ursachen zu ihren physischen Wirkungen führen. Manche physischen Wirkungen wären dann in eine Struktur *gemeinschaftlicher Verursachung mit nicht-physischen Teilursachen* eingebettet. In Abbildung 3.1 ist eine solche Kausalstruktur illustriert: Die physische Wirkung p* hat hier zwar eine physische Teilursache p. Allerdings tritt p* nur dann ein, wenn sowohl p als auch die nicht-physische Teilursache m eintreten. p allein ist also nicht hinreichend für p*. Ohne die nicht-physische Ursache m würde p* nicht eintreten. Das aber würde heißen, dass die Physik nicht – wie Kim es ausdrückt – kausal und explanatorisch eigenständig ist: Eine vollständige kausale Erklärung eines physischen Ereignisses müsste dann auch die nicht-physische Teilursache erwähnen.



Erläuterung: m ist eine nicht-physische Teilursache für p*. p ist eine physische Teilursache für p*. Die gestrichelten Pfeile stehen für die Relation der Teilverursachung.
p* tritt nur ein, wenn sowohl m als auch p eintreten. Weder m noch p sind für sich genommen hinreichend für p*.

Abbildung 3.1 Gemeinschaftliche Verursachung mit nicht-physischen Teilursachen

Die Kompatibilität mit Fällen gemeinschaftlicher Verursachung mit nicht-physischen Teilursachen kann durch folgende Formulierung vermieden werden:

(KG2) Alle physischen Ereignisse haben *hinreichende* physische Ursachen.[5]

Diese These ist insofern stärker als (KG1), als dass nicht bloß verlangt wird, dass physische Ereignisse unter anderem physische Teilursachen haben. Stattdessen wird verlangt, dass physische Ereignisse *hinreichende* physische Ursachen haben. Um eine vollständige Erklärung eines physischen Ereignisses zu geben, reicht es dann, auf dessen hinreichende physische Ursache zu verweisen. Eine vollständige Erklärung braucht entsprechend nicht auf nicht-physische Ursachen zu verweisen. Die Selbstständigkeit der Physik ist also durch (KG2) besser eingefangen als durch (KG1).

Dennoch ist (KG2) noch keine zufriedenstellende Formulierung der These der kausalen Geschlossenheit. Denn (KG2) ist noch immer kompatibel damit, dass die kausale Kette, die zu einem physischen Ereignis führt, als unverzichtbare Bestandteile nicht-physische Ursachen enthält. Dies ist insbesondere dann der Fall, wenn eine physische Wirkung in eine Struktur *gemischter Kausalketten mit nicht-physischen Gliedern* eingebettet ist.[6] Eine solche Struktur ist in Abbildung 3.2 illustriert: p* hat hier zwar eine hinreichende physische Ursache p. p verursacht p* jedoch nicht direkt, sondern indirekt über ein nicht-physisches, kausales Zwischenglied m. Eine Erklärung des Zustandekommens von p*, die dieses nicht-physische Zwischenglied unerwähnt lässt, wäre keine vollständige Erklärung. Wieder wäre also die Grundidee, dass vollständige Kausalerklärungen physischer Ereignisse nicht auf nicht-physische Ursachen verweisen müssen, nicht angemessen eingefangen.

Dieses Problem lässt sich durch eine der folgenden Modifikationen vermeiden:

(KG3) Alle physischen Ereignisse haben *direkte* hinreichende physische Ursachen.[7]

(KG4) Alle physischen Ereignisse haben hinreichende physische Ursachen *zu t* – wobei ‚*t*‘ für einen beliebigen Zeitpunkt vor Eintreten des Wirkungsereignisses steht.[8]

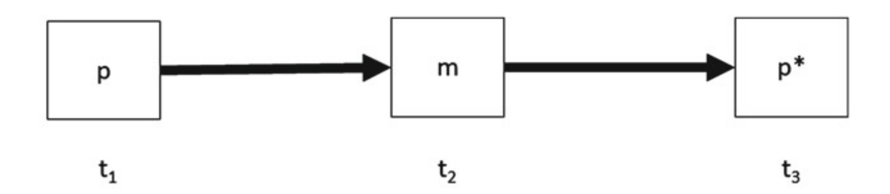
---

[5] Diese Formulierung ist als grobe Charakterisierung der These der kausalen Geschlossenheit sehr verbreitet. Siehe z. B. Papineau (2009: 53): „At first pass the causal closure of physics says that every physical effect has a sufficient physical cause.“

[6] Vgl. für eine ausführlichere Diskussion dieses Problems Lowe (2000).

[7] U.a. beruft sich Papineau (2009: 59) auf die Relation der direkten hinreichenden Verursachung: „Every physical effect has an immediate sufficient physical cause, in so far it has a sufficient physical cause at all.“

[8] Der Bezug auf Zeitpunkte wird u. a. von Kim (1989: 43) in Anspruch genommen: „Any physical event which has a cause at time t has a physical cause at t“.



Erläuterung: p und p* sind physische Ereignisse, m ist ein nicht-physisches Ereignis. Die dicken Pfeile stehen für die Relation der hinreichenden Verursachung.
p ist hinreichende Ursache für m und m ist hinreichende Ursache für p*. Unter der Annahme der Transitivität hinreichender Verursachung ist p also hinreichende Ursache für p*. Dennoch ist m unverzichtbarer Bestandteil einer vollständigen Kausalerklärung von p*.

Abbildung 3.2 Gemischte Kausalketten mit nicht-physischen Gliedern

These (KG3) schließt Fälle gemischter Kausalketten mit nicht-physischen Gliedern aus, indem sie festsetzt, dass physische Ereignisse *direkte* hinreichende physische Ursachen haben. Eine direkte hinreichende Ursache für p* ist dabei eine hinreichende Ursache, die p* ohne eine Vermittlung über kausale Zwischenglieder verursacht. In Abbildung 3.2 hat p* hingegen nur eine indirekte hinreichende physische Ursache, weshalb diese Situation gegen (KG3) verstößt. These (KG4) wählt einen anderen Weg: Gemischte Kausalketten mit nicht-physischen Gliedern sind laut dieser These ausgeschlossen, weil für jeden Zeitpunkt vor Eintreten des Wirkungsereignisses eine hinreichende physische Ursache existieren soll. In der in Abbildung 3.2 illustrierten Situation gibt es zu Zeitpunkt t_2 aber keine hinreichende physische Ursache, weshalb die Situation nicht mit (KG4) vereinbar ist. (KG3) und (KG4) sind also zwei unterschiedliche und gleichermaßen funktionierende Möglichkeiten, einer angemessenen Formulierung der These der kausalen Geschlossenheit näher zu kommen.[9]

Eine letzte Komplikation: Die bisherigen Formulierungen (KG2), (KG3) und (KG4) implizieren sämtlich, dass *alle* physischen Ereignisse hinreichende Ursachen haben. Diese Forderung steht jedoch mit der *Existenz indeterminierter*

[9] Lowe (2000: 576 ff.) wendet gegen These (KG4) ein, dass sie der Möglichkeit simultaner Verursachung nicht Rechnung trägt. Diese Komplikation ignoriere ich hier. Vgl. Garcia (2014, 98 ff.) für weitere Diskussion.

physischer Ereignisse in Konflikt, die durch die Quantenmechanik nahegelegt wird. Einige physische Ereignisse haben keine hinreichenden Ursachen – weder physische noch nicht-physische. Sie haben stattdessen lediglich eine bestimmte objektive Eintrittswahrscheinlichkeit. Üblicherweise wird die Existenz indeterminierter physischer Ereignisse jedoch nicht als Widerlegung der These der kausalen Geschlossenheit aufgefasst. Stattdessen wird die These der kausalen Geschlossenheit so formuliert, dass sie mit der Existenz indeterminierter physischer Ereignisse kompatibel ist.

Die im Kontext von Diskussionen mentaler Verursachung verbreitetste Formulierung ist die folgende:

> **(KG5)** Alle physischen Ereignisse, die eine hinreichende Ursache zu t haben, haben eine hinreichende physische Ursache zu t.

These (KG5) ist mit der Existenz indeterminierter physischer Ereignisse kompatibel, weil sie diese Ereignisse explizit ausklammert. Sie beschränkt sich auf determinierte physische Ereignisse und schweigt mit Blick auf die kausale Geschichte indeterminierter Ereignisse. Dies mag unbefriedigend wirken: Denn nun können die objektiven Eintrittswahrscheinlichkeiten indeterminierter physischer Ereignisse in Übereinstimmung mit (KG5) auf ganz beliebige Weise festgelegt sein. Insbesondere können sie ausschließlich durch nicht-physische Ereignisse festgelegt sein, so dass physische Ereignisse gar keine Rolle für das Zustandekommen indeterminierter physischer Ereignisse spielen. Auch in diesem Falle wäre die Physik nicht kausal selbstständig, da eine Kausalerklärung indeterminierter physischer Ereignisse auf nicht-physische Ereignisse verweisen müsste. Zudem droht (KG5) beinahe inhaltsleer zu werden, wenn sich die *meisten* physischen Ereignisse als indeterminiert erweisen. Tatsächlich ist (KG5) sogar damit kompatibel, dass nicht ein einziges physisches Ereignis tatsächlich eine hinreichende physische Ursache hat.

Diesen Bedenken lässt sich durch die folgende Formulierung begegnen:

> **(KG6)** Alle physischen Ereignisse haben hinreichende physische Ursachen zu t oder haben eine objektive Eintrittswahrscheinlichkeit, die durch andere physische Ereignisse festgelegt ist.

These (KG6) trifft auch eine Aussage über indeterminierte physische Ereignisse. Sie fängt die Grundidee, dass vollständige Kausalerklärungen physischer Ereignisse nicht auf nicht-physische Ereignisse verweisen müssen, daher angemessen ein.[10]

Die hier aufgeführten Formulierungen sprechen für sich genommen nicht gegen die These, dass nicht-physische Ereignisse gelegentlich physische Ereignisse verursachen. Sie verlangen, dass physische Ereignisse hinreichende physische Ursachen haben. Aber sie verlangen nicht, dass physische Ereignisse darüber hinaus keine nicht-physischen Ursachen haben. Diese Konsequenz wird im Rahmen des Vorwurfs des Epiphänomenalismus erst durch die Hinzunahme eines Exklusionsprinzip etabliert, das in Abschnitt 3.3 noch ausführlich besprochen wird.

Gelegentlich finden sich hingegen auch stärkere Formulierungen, die unmittelbar gegen die Verursachung physischer Ereignisse durch nicht-physische Ereignisse sprechen.[11] Im Kontext einer Diskussion des Vorwurfs des Epiphänomenalismus gegen den nicht-reduktiven Physikalismus sind solche Formulierungen jedoch ungünstig. Denn sie versperren den Blick auf die Relevanz der These der kausalen Exklusion. Die These der kausalen Exklusion ist es aber gerade, die von vielen nicht-reduktiven Physikalist*innen abgelehnt wird. Eine angemessene Darstellung der Dialektik erfordert es daher, die These der kausalen Exklusion als eigenständige Prämisse von der These der kausalen Geschlossenheit zu trennen und somit eine schwache Version der These der kausalen Geschlossenheit vorauszusetzen.

In diesem Abschnitt ist deutlich geworden, dass eine angemessene Formulierung der These der kausalen Geschlossenheit mit einer Reihe von Komplikationen

---

[10] Die Verbindung der These der kausalen Geschlossenheit zu kausalen Erklärungen, die auch Kim in Anspruch nimmt, hat in diesem Abschnitt eine veranschaulichende Funktion. Dies soll nicht darüber hinwegtäuschen, dass es sich bei der These der kausalen Geschlossenheit um eine metaphysische und nicht um eine epistemische These handelt: Sie betrifft in erster Linie die kausale Struktur der Welt und nicht (oder nur in einem abgeleiteten Sinne) unsere Erklärungen physischer Ereignisse.

[11] Montero (2003: 174ff) unterscheidet beispielsweise die hier diskutierten Versionen von einer ‚Strong Causal Closure'-These („Physical effects have only physical causes“) und einer ‚Two-Way Causal Closure'-These („There is no causal interaction between the physical and the non-physical'). Auch Kim beruft sich gelegentlich auf starke Versionen der These der kausalen Geschlossenheit. Vgl. z. B. Kim (2005: 50): „Any cause of a physical event is itself a physical event – that is, no non-physical event can be a cause of a physical event.“ (vgl. auch Kim (1998: 40). Auch er verweist aber darauf, dass die starke Version der These der kausalen Geschlossenheit dialektisch ungünstig ist (vgl. Kim (2005: 50–52)).

verbunden ist, denen durch verschiedene Qualifikationen Rechnung getragen werden kann. Wenn ich im Folgenden von der These der kausalen Geschlossenheit spreche, sollten diese Qualifikationen im Hinterkopf behalten werden. Ich werde die nötigen Qualifikationen jedoch nicht immer explizit machen. Stattdessen werde ich aus Gründen der einfacheren Darstellung häufig eine Formulierung wählen, die (KG2) recht nahe kommt: Physische Ereignisse haben hinreichende physische Ursachen. Dort, wo die Qualifikationen systematisch relevant werden, weise ich explizit darauf hin.

3.2.2 Argumente für die These der kausalen Geschlossenheit

Die These der kausalen Geschlossenheit ist eine sehr allgemeine, metaphysische These über die kausale Struktur der physischen Welt. Es stellt sich die Frage, wie diese allgemeine metaphysische These gerechtfertigt werden kann. Was also spricht dafür, dass der Bereich des Physischen im geschilderten Sinne kausal geschlossen ist?

Ein erstes, mögliches Argument, das in der zu Beginn des vorigen Abschnitts zitierten Bemerkung von Kim bereits angedeutet wird, beruft sich auf die prinzipielle Möglichkeit einer kausal-explanatorisch vollständigen Physik.[12] Kim wird an folgender Stelle noch etwas deutlicher:

> If you reject this principle, you are ipso facto rejecting the in-principle completability of physics – that is, the possibility of a complete and comprehensive physical theory of all physical phenomena. For you would be saying that any complete explanatory theory of the physical domain must invoke nonphysical causal agents. Never mind a complete physical explanation of everything there is; there couldn't even be a complete physical explanation of everything physical. It is safe to assume that no serious physicalist could accept such a prospect.[13]

Der Gedankengang in diesem Zitat lässt sich wie folgt zusammenfassen: Eine Ablehnung der These der kausalen Geschlossenheit würde bedeuten, dass die Physik aus prinzipiellen Gründen kausal-explanatorisch unvollständig bleiben muss. Dies wiederum würde heißen, dass nicht-physische Ursachen (wie irreduzible Akteure oder Götter) zur kausalen Erklärung physischer Ereignisse herangezogen werden müssen. Dies ist einerseits unplausibel, so dass sich ein

[12] Vgl. für eine Diskusssion des Arguments auch Hüttemann (2013b: 47–49) und Montero (2003: 177–179).

[13] Kim (1998: 40).

direktes Argument für die These der kausalen Geschlossenheit ergibt. Andererseits ist die notwendige Unvollständigkeit der Physik besonders unplausibel aus der Perspektive des Physikalismus, so dass sich ein (noch stärkeres) Argument dafür ergibt, dass (reduktive wie auch nicht-reduktive) Physikalist*innen auf die These der kausalen Geschlossenheit festgelegt sind.

Die auf diese Weise angedeuteten Argumente können zumindest an zwei Stellen angegriffen werden.

Erstens folgt aus der These, dass die Physik notwendig unvollständig ist, keinesfalls, dass es vollständige Erklärungen gibt, die auf nicht-physische Entitäten wie Götter oder irreduzible Akteure verweisen. Dass eine physikalische Erklärung unmöglich ist, bedeutet nicht unbedingt, dass eine nicht-physikalische Erklärung möglich ist. Die in Frage stehenden Ereignisse könnten auch prinzipiell unerklärbar sein. Somit kann der Verweis auf unplausible Alternativerklärungen hier keine argumentative Arbeit leisten.

Zweitens kann in Zweifel gezogen werden, dass die Physik im relevanten Sinne prinzipiell vervollständigbar ist. Dies ist schließlich eine nicht-triviale Annahme, die weiterer Rechtfertigung bedarf.[14] Diese Reaktion scheint jedoch aus dualistischer Perspektive attraktiver als aus der Perspektive des (nicht-reduktiven) Physikalismus. Es besteht zumindest prima facie eine Spannung zwischen der These, dass der Bereich des Physischen die metaphysische Grundlage für alle Entitäten bildet, und der Idee, dass die Physik prinzipiell unvollständig ist. Wenn diese Spannung sich als belastbar erweist, wäre also zumindest Kims Argument dafür, dass (nicht-reduktive) Physikalist*innen auf die These der kausalen Geschlossenheit festgelegt sind, an dieser Stelle überzeugend.

Ein zweiter Ansatz zur Begründung der These der kausalen Geschlossenheit beruht auf einem Induktionsschluss. Die Grundidee bringt Andrew Melnyk auf den Punkt:

> [C]urrent physics‘ success to date in finding that many physical events have sufficient physical causes provides inductive evidence that all physical events, including both unexamined physical events and examined-but-as-yet-unexplained physical events, have sufficient physical causes.[15]

---

[14] Hüttemann (2013b: 48) weist in Zusammenhang mit der zitierten Bemerkung von Kim darauf hin, dass „gar nicht einzusehen ist, dass man die Möglichkeit einer explanatorisch vollständigen und umfassenden physikalischen Theorie unabhängig von der Erfahrung akzeptieren muss, um z. B. erfolgreich Physik zu betreiben“, und „dass es letztlich eine Angelegenheit ist, die sich im Verlaufe der Wissenschaftsgeschichte zeigt, ob sich eine solche Theorie tatsächlich konstruieren lässt.“

[15] Melnyk (2003b: 161).

Daraus, dass viele physische Ereignisse nach unserem jetzigen Wissensstand hinreichende physische Ursachen haben, können wir demzufolge induktiv schließen, dass alle physischen Ereignisse hinreichende physische Ursachen haben.

Wie überzeugend ist dieses induktive Argument für die These der kausalen Geschlossenheit?

Erstens stellt sich die Frage nach der Stärke der Induktionsbasis. Tatsächlich schneidet das induktive Argument für die These der kausalen Geschlossenheit hier auf den ersten Blick gut ab: Die kausalen Erklärungserfolge der Physik sind weitreichend. Es handelt sich nicht um einige wenige, isolierte Erfolge. Zudem war es bisher nie nötig, auf nicht-physische Ursachen zu verweisen, um diese Erklärungsfolge zu leisten.[16]

Zweitens könnte man bezweifeln, dass die Klasse der physischen Ereignisse hinreichend homogen ist, um einen Induktionsschluss zu erlauben. Ließen sich relevante Unterschiede ausmachen zwischen den bisher erklärten physischen Ereignissen und den unerklärten physischen Ereignissen, auf deren kausale Erklärbarkeit geschlossen wird, würde der Induktionsschluss dadurch an Überzeugungskraft verlieren. Auch hier scheint es jedoch keine starken Gründe zu geben, einen solchen Unterschied zu erwarten. Aus wissenschaftlicher Perspektive sind z. B. die physischen Ereignisse in menschlichen Gehirnen, oder die mit Handlungen verbundenen Körperbewegungen, nicht besonders gegenüber anderen physischen Ereignissen. Warum also sollte man davon ausgehen, dass diese Ereignisse prinzipiell unerklärbar durch andere physische Ereignisse sind?[17]

Das induktive Argument zeigt also, dass es zumindest starke Evidenzen für die These der kausalen Geschlossenheit gibt. Natürlich bleibt Raum für Zweifel: Zukünftige Entwicklungen in der Physik oder Physiologie könnten zu der Überzeugung führen, dass manche physischen Ereignisse keine hinreichenden physischen Ursachen haben. Es gibt jedoch keinen Grund, solche Entwicklungen zu erwarten.

Ein drittes Argument nimmt seinen Ausgang vom Energieerhaltungssatz in der modernen Physik. In geschlossenen Systemen bleibt die Gesamtenergie konstant:

---

[16] Vgl. Vicente (2006: 155ff) für weitere, kritische Diskussion. Widerspruch ist insbesondere von Emergentist*innen zu erwarten, die dafür argumentieren, dass bereits aus Sicht der heutigen Physik einige physische Ereignisse nicht erklärt werden können, ohne auf nicht-physische Ereignisse Bezug zu nehmen.

[17] Vgl. Melnyk (2003: 161f). Papineaus (2001) ‚Argument aus der Physiologie' bezieht sich zudem explizit auf Erklärungserfolge der Physiologie mit Blick auf physische Ereignisse, die üblicherweise psychologisch erklärt werden. Hieraus ergibt sich also ein induktives Argument, das spezifischer auf solche physischen Ereignisse bezogen ist.

Es geht weder Energie verloren, noch entsteht neue Energie. Das physische Universum wird in diesem Argument als geschlossenes System aufgefasst. Daher ist durch den Energieerhaltungssatz ausgeschlossen, dass dem physischen Universum ‚von außerhalb' Energie zugeführt wird, oder dass Energie aus dem physischen Universum ‚nach außen' abgegeben wird.

Ohne auf die Details dieser Argumentation einzugehen, lassen sich zwei Aspekte dieser Idee hervorheben:

Erstens: Um vom Energieerhaltungssatz zu einer gehaltvollen These der kausalen Geschlossenheit zu gelangen, muss Kausalität mit Energieübertragung verknüpft werden. Wenn Verursachung notwendig mit Energieübertragung einherginge, könnte tatsächlich keine nicht-physische Ursache eine physische Wirkung haben, ohne dass der Energieerhaltungssatz verletzt wird. Ob eine solche Verbindung zwischen Energieübertragung und Kausalität aber hergestellt werden kann, ist unklar. Hier hängt einiges von der vorausgesetzten Theorie der Kausalität ab. Wie wir (in Kapitel 4) sehen werden, folgt ein solcher Zusammenhang unmittelbar aus z. B. Phil Dowes Erhaltungsgrößentheorie der Kausalität.[18] Im Rahmen einer kontrafaktischen oder interventionistischen Theorie ist ein solcher Zusammenhang aber nicht ohne Weiteres zu begründen.[19]

Zweitens: Die Argumentation scheint im Erfolgsfall nicht nur zu etablieren, dass physische Ereignisse hinreichende physische Ursachen haben. Vielmehr scheint es zusätzlich zu begründen, dass physische Ereignisse keine nicht-physischen Ursachen haben. Es wird also eine Geschlossenheitsthese motiviert, die stärker ist als die Formulierungen, die ich in Abschnitt 3.1.1. eingeführt habe. Dies hat zum einen zur Konsequenz, dass nicht-reduktive Physikalist*innen, die an einer Verteidigung mentaler Verursachung interessiert sind, schlecht darin beraten wären, sich auf dieses Argument für die These der kausalen Geschlossenheit zu berufen.[20] Zum anderen sollten aber auch reduktive Physikalist*innen, die den nicht-reduktiven Physikalismus auf Grundlage der These der kausalen Geschlossenheit angreifen wollen, sich nicht auf dieses Argument berufen. Denn ein Angriff auf den nicht-reduktiven Physikalismus auf Grundlage einer starken Geschlossenheitsthese gilt gemeinhin als dialektisch ungünstig.[21]

[18] Vgl. Dowe (2000).

[19] Vgl. auch die Diskussion in Vicente (2006: 162ff).

[20] Vgl. Gibb (2012).

[21] Vgl. hierzu z. B. Kim (2005: 50–52).

3.2.3 Die Reichweite der These der kausalen Geschlossenheit

Die These der kausalen Geschlossenheit besagt, dass der Bereich *des Physischen* kausal geschlossen ist. Im Lichte der Unterscheidungen zwischen verschiedenen Sinnen von ‚physisch', die ich in den Abschnitt 2.3.2 und 2.3.3. getroffen habe, bedarf diese These jedoch weiterer Spezifikation. Sie kann sich zum einen auf physische$_{weit}$ Ereignisse beziehen, so dass auch chemische oder biologische Ereignisse betroffen sind. Zum anderen kann sie sich auf physische$_{eng}$ Ereignisse beziehen, so dass in erster Linie Ereignisse betroffen sind, die Gegenstand der Physik sind. Welche Spezifikationen der These der kausalen Geschlossenheit sind also plausibel?

Eine Komplikation bei der Beantwortung dieser Frage ergibt sich daraus, dass der Ausdruck ‚physisch' in der These der kausalen Geschlossenheit zweimal vorkommt: *Physische* Ereignisse haben hinreichende *physische* Ursachen. Es gibt daher die Möglichkeit, die beiden Vorkommnisse des Ausdrucks unterschiedlich zu deuten. Es ergeben sich somit die folgenden vier möglichen Formulierungen:[22]

(KG7) Alle physischen$_{eng}$ Ereignisse haben hinreichende physische$_{eng}$ Ursachen.

(KG8) Alle physischen$_{weit}$ Ereignisse haben hinreichende physische$_{weit}$ Ursachen.

(KG9) Alle physischen$_{eng}$ Ereignisse haben hinreichende physische$_{weit}$ Ursachen.

(KG10) Alle physischen$_{weit}$ Ereignisse haben hinreichende physische$_{eng}$ Ursachen.

Unter der Voraussetzung, dass die Menge der physischen$_{eng}$ Ereignisse eine echte Teilmenge der Menge der physischen$_{weit}$ Ereignisse ist, ist These (KG10) dabei die logisch stärkste These: Sie impliziert die Thesen (KG7), (KG8) und (KG9), ohne von einer dieser Thesen impliziert zu werden. These (KG9) ist hingegen die logisch schwächste These: Sie wird von jeder der anderen Thesen impliziert, ohne eine der anderen Thesen zu implizieren. These (KG9) wäre zum Beispiel damit vereinbar, dass manche mikrophysikalischen Ereignisse keine hinreichende mikrophysikalischen Ursachen haben, sondern ausschließlich hinreichende biologische Ursachen. Auch These (KG8) wäre mit dieser Situation kompatibel.

Zwei Überlegungen sprechen dafür, dass These (KG7) der primäre Sinn der These der kausalen Geschlossenheit ist:

---

[22] Ich verwende hier als Ausgangspunkt die simple Formulierung (KG2). Die nötigen Qualifikationen, die ich in Abschnitt 3.2.1. ausgeführt habe, lasse ich implizit.

Erstens: Die in Abschnitt 3.2.2 diskutierten Argumente für die These der kausalen Geschlossenheit scheinen in erster Linie für These (KG7) zu sprechen. Kim führt die prinzipielle kausal-explanatorische Vervollständigbarkeit *der Physik* als Ausgangspunkt seines Arguments für die These der kausalen Geschlossenheit ins Feld. Aus der Vervollständigbarkeit der Physik folgt jedoch nicht unmittelbar etwas für die kausale Geschichte von z. B. biologischen Ereignissen. Auch das induktive Argument in der zitierten Version von Melnyk bezieht sich auf den vergangenen Erklärungserfolg *der Physik*. Es ist daher zumindest naheliegend, auch dieses Argument in erster Linie auf jene Ereignisse zu beziehen, die in der Physik behandelt werden.[23] Auch jene Argumente, die vom Grundsatz der Energieerhaltung ausgehen, betreffen zunächst ausschließlich $\text{physische}_{\text{eng}}$ Ereignisse. Es scheint also, dass primär These (KG7) durch die besprochenen Argumente begründet ist.

Da These (KG7) These (KG9) impliziert, weiten sich die Argumente auch auf diese These aus: Wenn alle $\text{physischen}_{\text{eng}}$ Ereignisse hinreichende $\text{physische}_{\text{eng}}$ Ursachen haben, dann haben auch alle $\text{physischen}_{\text{eng}}$ Ereignisse $\text{physische}_{\text{weit}}$ Ursachen – einfach deshalb, weil die Menge der $\text{physischen}_{\text{eng}}$ Ereignisse eine Teilmenge der Menge der $\text{physischen}_{\text{weit}}$ Ereignisse ist. These (KG9) ist dementsprechend begründet, *weil* These (KG7) begründet ist. Die Verursachung mikrophysikalischer Ereignisse ausschließlich durch biologische oder andere höherstufige, $\text{physische}_{\text{weit}}$ Ereignisse, die von (KG9) eigentlich zugelassen wird, ist vor dem Hintergrund dieser Begründung aber ausgeschlossen.

Zweitens: Plausiblerweise ist neben dem Bereich des im engen Sinne Physischen kein weiterer Bereich kausal geschlossen. Beispielsweise haben nicht alle biologischen Ereignisse hinreichende biologische Ursachen. Papineau gibt hierfür das Beispiel des Aussterbens der Dinosaurier – einem biologischen Ereignis – das durch einen Kometeneinschlag verursacht wurde – einem nicht-biologischen Ereignis.[24] Im Allgemeinen ist es unwahrscheinlich, dass wir für alle höherstufigen, $\text{physischen}_{\text{weit}}$ Ereignisse andere höherstufige, $\text{physische}_{\text{weit}}$ Ursachen finden können: Die Entstehung von Higgs-Bosonen im LHC – ein $\text{physisches}_{\text{eng}}$ Ereignis – hatte zahlreiche $\text{physische}_{\text{weit}}$ Wirkungen, für die wir keine hinreichende $\text{physischen}_{\text{weit}}$ Ursachen finden. Sie verursachte zum Beispiel – so können wir

[23] Hier ist die Sache jedoch nicht ganz eindeutig: Erstens lassen sich auch induktive Argumente vorbringen, die den Erklärungserfolg der Naturwissenschaften im weiteren Sinne als Induktionsbasis nehmen. Zweitens könnte man argumentieren, dass der Umstand, dass viele $\text{physischen}_{\text{eng}}$ Ereignisse hinreichende $\text{physische}_{\text{eng}}$ Ursachen haben, einen induktiven Schluss auf alle $\text{physischen}_{\text{weit}}$ Ereignisse erlaubt.

[24] Vgl. Papineau (2009: 54).

annehmen – die Ausschüttung von Endorphinen in den Gehirnen zahlreicher Wissenschaftler – also neurologische Ereignisse. Diese Ereignisse können wir nicht vollständig durch Verweis auf ausschließlich höherstufige, $\text{physische}_{\text{weit}}$ Ursachen erklären. Der Bereich der höherstufigen, $\text{physischen}_{\text{weit}}$ Ereignisse ist also *nicht* kausal geschlossen.[25]

Diese Überlegung zeigt, dass wir die hinreichenden Ursachen aller $\text{physischen}_{\text{weit}}$ Ereignisse nicht ausschließlich aus dem Bereich der *höherstufigen* $\text{physischen}_{\text{weit}}$ Ereignisse ziehen können. Sie ist aber damit vereinbar, dass wir für alle $\text{physischen}_{\text{weit}}$ Ereignisse hinreichende *physische*$_{eng}$ Ursachen finden können. Sie ist also mit These (KG10) kompatibel. In Abschnitt 3.3.7. werde ich dafür argumentieren, dass auch These (KG10) unter Voraussetzung weiterer Annahmen – insbesondere des Prinzips der Aufwärtsverursachung und der These der ontologischen Abhängigkeit – aus These (KG7) folgt. Das hieße, dass alle vier Thesen begründet werden können. Es ist jedoch wichtig, dass eine solche Argumentation auf einem zusätzlichen und möglicherweise kontroversen Prinzip beruht. Erst einmal ist nach meiner Auffassung ausschließlich These (KG7) begründet: Die Reichweite der These der kausalen Geschlossenheit beschränkt sich auf $\text{physische}_{\text{eng}}$ Ereignisse.

3.2.4 Kausalität und kausale Geschlossenheit

Die These der kausalen Geschlossenheit ist eine *kausale* These. Sie sagt also etwas darüber aus, in welchen *kausalen* Verhältnissen physische Ereignisse stehen. Dies wirft die Frage auf, was hier genauer mit Kausalität gemeint ist. Ausführlich auf die Relation der Kausalität werde ich erst in Kapitel 4 eingehen. An dieser Stelle möchte ich jedoch bereits auf einige implizite Voraussetzungen der These der kausalen Geschlossenheit eingehen, die den Begriff der Kausalität betreffen.

Erstens setzt die These der kausalen Geschlossenheit voraus, dass es einen sinnvollen Begriff von *hinreichender* Verursachung gibt. Wie in Abschnitt 3.2.1. bereits angedeutet, ist die Relation der hinreichenden Verursachung dabei nicht

[25] Vgl. auch Kim (2005: 65): „It is only when we reach the fundamental level of microphysics that we are likely to get a causally closed domain.“; Baker (1993: 79): „Since the notion of causal closure of the physical applies only on the lowest level of micro-physics, the only sense of ‘physical’ in which the causal closure of the physical has a chance of being true is micro-physical, where ‘micro-physical’ is a name for whatever turn out to be basic physical particles and their properties.“

mit der Relation der Verursachung (simpliciter) gleichzusetzen: Nicht alle Ursachen sind hinreichende Ursachen. Um ein verbreitetes Beispiel[26] zu bemühen: Ein Kurzschluss kann einen Hausbrand verursachen, ohne dass der Kurzschluss für sich genommen hinreichend für den Hausbrand ist. Erst gemeinsam mit vielen anderen Faktoren – wie der Anwesenheit von brennbarem Material und dem Umstand, dass die üblichen Brandschutzmechanismen nicht funktionieren – bildet der Kurzschluss (eventuell) eine hinreichende Bedingung für den Brand. Der Kurzschluss ist also eine Ursache für den Brand, aber keine *hinreichende* Ursache.[27]

Anhand dieses Beispiels lässt sich zudem veranschaulichen, dass hinreichende Bedingungen für ein Ereignis sehr komplex sein müssen. Hierfür muss man sich vor Augen führen, was zu einem Kurzschluss alles hinzukommen muss, damit ein Hausbrand garantiert ist: Einerseits müssen zahlreiche ‚positive' Bedingungen wie die Anwesenheit von brennbarem Material oder die Anwesenheit von Sauerstoff erfüllt sein. Andererseits müssen aber auch zahlreiche ‚negative' Bedingungen gegeben sein: Es darf *nicht* der Fall sein, dass das brennbare Material (aus welchen Gründen auch immer) nass ist; es darf *nicht* der Fall sein, dass ein Bewohner den ersten Funken bemerkt und löscht, usw. Allgemein darf gewissermaßen nichts dazwischenkommen, was den Brand trotz des Kurzschlusses verhindern würde.[28]

---

[26] Vgl. Mackie (1965).

[27] Gelegentlich wird auch ein schwächerer Begriff von hinreichender Verursachung verwendet, demzufolge hinreichende Ursachen auch Ursachen sein können, die *in den gegebenen Umständen* hinreichend sind. In diesem Sinne wäre der Kurzschluss möglicherweise als hinreichende Ursache zu bezeichnen. Hinreichende Verursachung kann im Kontext der These der kausalen Geschlossenheit jedoch nicht im Sinne von ‚hinreichend-in-den-gegebenen-Umständen' interpretiert werden, weil eine so interpretierte These der kausalen Geschlossenheit damit vereinbar wäre, dass die Umstände teilweise nicht-physisch sind. Das oben angeführte Problem der Fälle von gemeinschaftlicher Verursachung mit nicht-physischen Teilursachen wäre so also nicht in den Griff zu kriegen: P in Abbildung 3.1 ist *in den gegebenen Umständen* hinreichende Ursache für P*, wenn die gegebenen Umstände das Stattfinden von M beinhalten.

[28] Mumford und Anjum (2011: 56 ff.) präsentieren ein Argument dafür, dass Theorien der Kausalität, die negative Faktoren (wie die Abwesenheit von verhindernden Faktoren) nicht als Ursachen klassifizieren, daher dazu führen, dass es im Allgemeinen keine hinreichenden Ursachen gibt: Die Ursachen eines Ereignisses sind nie hinreichend für das Eintreten eines Wirkungsereignisses, da sie in Anwesenheit eines verhinderenden Faktors nicht zu dem Wirkungsereignis geführt hätten. Hinreichende Bedingungen müssen also immer die Abwesenheit verhindernder Faktoren beinhalten und können daher nicht vollständig aus Ursachen zusammengesetzt sein. Vgl. Gibb (2015: 631 ff.) für ein verwandtes Argument.

Die in Kapitel 4 zu diskutierenden Theorien der Kausalität stellen nicht den Begriff der hinreichenden Verursachung in den Mittelpunkt, sondern den Begriff der Verursachung (simpliciter). Es ist nicht trivial, innerhalb dieser Theorien einen angemessenen Begriff der hinreichenden Verursachung zu entwickeln. Vor diesem Hintergrund vertreten einige Autor*innen einen prinzipiellen Skeptizismus hinsichtlich des Begriffs der hinreichenden Verursachung. Panu Raatikainen schreibt etwa besonders deutlich:

> There are causes, which are difference-makers; and there are sufficient conditions, which are wholly different issues and not causes of any sort; there are no such things as sufficient causes.[29]

Setzt man einen solchen Skeptizismus hinsichtlich des Begriffs der hinreichenden Verursachung voraus, stellt dies die These der kausalen Geschlossenheit in Frage. Der Gehalt der These bleibt dann unklar, weil der zentrale Begriff der hinreichenden Verursachung nicht sinnvoll expliziert werden kann. Es bleibt also vorerst festzuhalten, dass die These der kausalen Geschlossenheit voraussetzt, dass der Begriff der hinreichenden Verursachung sinnvoll ist.

Eine zweite Voraussetzung der These der kausalen Geschlossenheit betrifft die Frage, ob es im Bereich des (im engen Sinne) Physischen überhaupt Ursachen gibt. Einige Autor*innen bezweifeln, dass der Begriff der Verursachung sinnvoll auf die Physik anwendbar ist.[30] Ein möglicher Grund für eine solche Skepsis ist, dass sich die Asymmetrie der Verursachungsrelation in den Gesetzen physikalischer Theorien nicht wiederfindet. Sollte es keine Verursachung im Bereich der Physik geben, spricht das ebenfalls gegen die These der kausalen Geschlossenheit. Denn physische$_{\text{eng}}$ Ereignisse hätten dann im allgemeinen keine physischen$_{\text{eng}}$ Ursachen. Entsprechend hätten sie auch keine *hinreichenden* physischen$_{\text{eng}}$ Ursachen. Die These der kausalen Geschlossenheit – zumindest in der Version, die sich auf physische$_{\text{eng}}$ Ereignisse bezieht[31] – setzt also voraus,

[29] Raatikainen (2010: 360). Ähnliche Tendenzen finden sich u. a. bei Menzies (2015: 27–30). Vgl. auch Raatikainen (2018: 33–40).

[30] Die Klassiker dieser Position sind Mach (1896) und Russell (1912). Für weitere Diskussion siehe Price und Corry (2007). Vgl. auch Hüttemann (2013a: 46–53). Eine ausführliche Verteidigung der Anwendbarkeit des Kausalitätsbegriffs in der Physik bietet Frisch (2014).

[31] Die Option, als Reaktion hierauf zu einer These der kausalen Geschlossenheit zu wechseln, die sich auf physische$_{\text{weit}}$ Ereignisse bezieht und dabei physische$_{\text{eng}}$ Ereignisse explizit ausschließt, diskutiert Ney (2016b: 148–153).

dass der Skeptizismus mit Blick auf Verursachung in der Physik zurückgewiesen werden kann.[32]

Beide genannten Voraussetzungen der These der kausalen Geschlossenheit zum Begriff der Verursachung können umgangen werden, indem die These der kausalen Geschlossenheit ganz von ihrem *kausalen* Gehalt befreit wird. Tatsächlich finden sich in der Literatur auch Varianten der These der Geschlossenheit, die auf kausales Vokabular verzichten. Man betrachte etwa die folgende *These der nomologischen Geschlossenheit*:

> **Nomologische Geschlossenheit (NG):** Die objektive Eintrittswahrscheinlichkeit aller physischen Ereignisse ist durch physische Gesetze und andere (frühere) physische Ereignisse vollständig festgelegt.[33]

Eine solche These der nomologischen Geschlossenheit setzt weder voraus, dass es einen sinnvollen Begriff von hinreichender Verursachung gibt, noch setzt sie voraus, dass es Verursachung im Bereich der Physik gibt. Wer also skeptisch gegenüber einer dieser Voraussetzungen der These der kausalen Geschlossenheit ist, kann noch immer die verwandte These der nomologischen Geschlossenheit akzeptieren.[34]

3.2.5 Die These der kausalen Geschlossenheit im nicht-reduktiven Physikalismus

Die These der kausalen Geschlossenheit wird oft als ein wesentlicher Bestandteil des (reduktiven und nicht-reduktiven) Physikalismus behandelt. Insbesondere wird der nicht-reduktive Physikalismus häufig so *definiert*, dass er auf die These der kausalen Geschlossenheit festgelegt ist. Meine Charakterisierung

[32] Weitere Diskussion findet sich bei Ney (2016b) und Haug (2019).

[33] Vgl. z. B. Papineau (2001: 9): „All physical effects are fully determined by law by prior physical occurrences“. Vgl. auch Loewer (2015: 41): „For every physical event proposition E(t’) (except events at the initial condition) and every prior time t, the probability at t of E(t’) given the physical state of the universe at t and the fundamental laws is the objective probability at of E(t’)“. Vgl. auch Ney (2016b: 153–159).

[34] Es stellt sich jedoch die Frage, inwieweit die Argumente für die These der kausalen Geschlossenheit auch die These der nomologischen Geschlossenheit stützen. Zudem ist fragwürdig, ob die These der nomologischen Geschlossenheit die argumentative Rolle in Argumenten für den Physikalismus oder gegen den nicht-reduktiven Physikalismus übernehmen kann, die üblicherweise von der These der kausalen Geschlossenheit übernommen wird. Für weitere Diskussion vgl. Ney (2016b).

des nicht-reduktiven Physikalismus in Teil 2 erwähnt die These der kausalen Geschlossenheit jedoch nicht und weicht in dieser Hinsicht also von anderen, weit verbreiteten Charakterisierungen ab.

Ein Grund hierfür ist, wie bereits angedeutet, dass der Kern des nicht-reduktiven Physikalismus für die Zwecke dieser Arbeit weitgehend von kausalen Konnotationen freigehalten werden sollte. Die Leitfrage dieser Arbeit lautet, ob der nicht-reduktive Physikalismus – verstanden als die Kombination aus den Thesen des Realismus, der Nicht-Identität und der ontologischen Abhängigkeit – gegen den Vorwurf des Epiphänomenalismus verteidigt werden kann. Für die Diskussion dieser Leitfrage ist es wichtig, zu verdeutlichen, dass diese Kombination von Thesen zunächst einmal – d. h. ohne weitere Argumente – keine Implikationen für das Bestehen oder Nicht-Bestehen irgendwelcher Kausalbeziehungen hat.

Dies kann vor dem Hintergrund der Überlegungen in Abschnitt 3.2.4. noch etwas deutlicher gemacht werden: Es spricht aus der Sicht des nicht-reduktiven Physikalismus zunächst einmal nichts dagegen, einen Skeptizismus mit Blick auf den Begriff der hinreichenden Verursachung oder mit Blick auf Kausalität in der Physik zu vertreten.[35] Diese Positionen implizieren jedoch, dass die These der kausalen Geschlossenheit falsch ist. Dann aber spricht aus der Sicht des nicht-reduktiven Physikalismus auch nichts gegen die Ablehnung der These der kausalen Geschlossenheit.

Diese Überlegungen zeigen jedoch bestenfalls, dass nicht-reduktive Physikalist*innen nicht auf die These der *kausalen* Geschlossenheit festgelegt sind. Sie sind vereinbar damit, dass nicht-reduktive Physikalist*innen auf die These der *nomologischen* Geschlossenheit festgelegt sind. Auch sind sie damit vereinbar, dass nicht-reduktive Physikalist*innen auf die These der kausalen Geschlossenheit festgelegt sind, wenn bestimmte weitere Bedingungen erfüllt sind. Zu diesen Bedingungen gehören zum Beispiel die Falschheit des Skeptizismus mit Blick auf den Begriff der hinreichenden Verursachung und mit Blick auf Kausalität in der Physik. Es könnte also sein, dass nicht-reduktive Physikalist*innen auf die These der kausalen Geschlossenheit festgelegt sind, *wenn sie diese Positionen ablehnen.*

In diesem eingeschränkten Sinn halte ich es für durchaus plausibel, dass nicht-reduktive Physikalist*innen auf die These der kausalen Geschlossenheit festgelegt sind: Sie sind auf die These der kausalen Geschlossenheit festgelegt, insofern sie den Begriff der hinreichenden Verursachung für sinnvoll halten und einen Platz für Verursachung in der Physik sehen. Wenn sie eine dieser Voraussetzungen

[35] Ebenso ist ein universeller Kausalitätsskeptizismus zunächst mit dem nicht-reduktiven Physikalismus kompatibel.

ablehnen, sind sie noch immer auf die These der nomologischen Geschlossenheit festgelegt. Denn würden sie diese These ablehnen, müssten sie tatsächlich die prinzipielle Vervollständigbarkeit der Physik ablehnen.

3.3 Vom nicht-reduktiven Physikalismus zum Epiphänomenalismus

In diesem Abschnitt erkunde ich einige Möglichkeiten, die erste Prämisse des Vorwurfs des Epiphänomenalismus zu begründen. Es geht also um die folgende These:

> **(VE1)** Wenn der nicht-reduktive Physikalismus wahr ist, dann gibt es keine mentale Verursachung.

Wenn diese These stimmt, werden nicht-reduktive Physikalist*innen in den Epiphänomenalismus gedrängt. Die einzige Möglichkeit einer Verteidigung des nicht-reduktiven Physikalismus bestünde dann in einer epiphänomenalistischen Erwiderung, die die Existenz mentaler Verursachung abstreitet. Wie also versuchen Kritiker*innen des nicht-reduktiven Physikalismus die These zu begründen?

Zentral für die Begründung der These sind *Exklusionsargumente*.[36] Die Grundidee von Exklusionsargumenten bringt Jaegwon Kim wie folgt knapp auf den Punkt:

> [T]he problem of causal exclusion is to answer this question: *Given that every physical event that has a cause has a physical cause, how is a mental cause also possible?*[37]

[36] Außer Acht lasse ich hier verschiedene Argumente, die den Epiphänomenalismus unabhängig vom nicht-reduktiven Physikalismus begründen. Zu nennen wären hier zum Beispiel das Argument aus dem semantischen Externalismus (vgl. z. B. Jacob (1997) für Diskussion) sowie das Argument aus der Anomalie des Mentalen (vgl. Davidson (1970) und Honderich (1982)). Diese Argumente nehmen ihren Ausgang nicht direkt von den Kernthesen des nicht-reduktiven Physikalismus.

Trivialerweise ist das Konditional ‚Wenn der nicht-reduktive Physikalismus wahr ist, dann gibt es keine mentale Verursachung' auch dann wahr, wenn es aus unabhängigen Gründen keine mentale Verursachung gibt. Mir geht es hier aber natürlich um Argumente, die zeigen, dass der nicht-reduktive Physikalismus auch relevant dafür ist, dass es keine mentale Verursachung gibt. Und hierfür sind in erster Linie Exklusionsargumente zu besprechen.

[37] Kim (1998: 38).

Kim spielt in dieser Frage zunächst auf die These der kausalen Geschlossenheit an, die ich in Abschnitt 3.2. bereits ausführlich diskutiert habe. Diese stellt sicher, dass jedes physische Ereignis, das als Wirkung eines mentalen Ereignisses in Frage kommt, bereits eine hinreichende physische Ursache hat. Wenn die (vorgeblichen) Wirkungen mentaler Ereignisse aber bereits hinreichende physische Ursachen haben, bleibt die kausale Rolle distinkter, mentaler Ereignisse rätselhaft. Mentale Ereignisse könnten ihre physischen Wirkungen allenfalls überdeterminieren. Diese Option hält Kim jedoch für überaus unplausibel.

Diese Grundidee von Exklusionsargumenten kann auf unterschiedliche Weisen entwickelt und erweitert werden. Im Folgenden gehe ich auf eine Reihe von möglichen Formulierungen von Exklusionsargumenten ein. Dabei gehe ich wie folgt vor: In Abschnitt 3.3.1. rekonstruiere ich zunächst ein sehr einfaches Exklusionsargument. In Abschnitt 3.3.2. zeige ich auf, dass diese einfache Version eine deutlich eingeschränkte Reichweite hat und deshalb nicht überzeugen kann. In Abschnitt 3.3.3. bespreche ich eine konservative Änderung des einfachen Exklusionsarguments, die die Reichweite des Arguments möglicherweise ausweitet. In Abschnitt 3.3.4. führe ich eine wichtige begriffliche Grundlage für die weitere Diskussion von Exklusionsargumenten ein, indem ich mentale Verursachung in einem Stufenmodell der Realität verorte. Darauf aufbauend führe ich in Abschnitt 3.3.5. drei Prinzipien ein, die den Zusammenhang von Verursachung und ontologischer Abhängigkeit betreffen. Diese drei Prinzipien eröffnen drei verschiedene Möglichkeiten, das einfache Exklusionsargument zu erweitern oder zu reformulieren, um seine Reichweite deutlich auszuweiten. Diese Möglichkeiten entwickle ich in den Abschnitten 3.3.6., 3.3.7. und 3.3.8.

In Abschnitt 3.3.9. schließlich gehe ich genauer auf das Exklusionsprinzip ein, das das Herzstück aller bis dahin besprochenen Exklusionsargumente bildet. Diese Diskussion führt zu einer Modifikation der besprochenen Exklusionsargumente, die ich in Abschnitt 3.3.10 einführe. In Abschnitt 3.3.11. stelle ich die besprochenen Exklusionsargumente abschließend noch einmal in einen Bezug zum Vorwurf des Epiphänomenalismus.

3.3.1 Das einfache Exklusionsargument

Das einfache Exklusionsargument nimmt seinen Ausgang von den folgenden beiden Thesen:

> **These der Nicht-Identität**: Mentale Ereignisse sind nicht mit physischen Ereignissen identisch.[38]
>
> **These der kausalen Geschlossenheit**: Physische Ereignisse, die hinreichende Ursachen zu t haben, haben hinreichende physische Ursachen zu t.[39]

Beide These werden in der Regel von nicht-reduktiven Physikalist*innen akzeptiert.[40] Sie können im vorliegenden Kontext als gegeben vorausgesetzt werden. Es ist also die Kombination dieser beiden Thesen, die laut dem einfachen Exklusionsargument dazu führt, dass der nicht-reduktive Physikalismus ein Problem mit mentaler Verursachung hat.

Das einfache Exklusionsargument lässt sich am besten anhand einer Anwendung dieser beiden Thesen auf ein schemenhaftes Szenario entwickeln. Gehen wir also davon aus, dass p* ein physisches Ereignis ist, das als Wirkung eines mentalen Ereignisses m (zum Beispiel Susis Wunsch) in Frage kommt. Laut der These der kausalen Geschlossenheit können wir postulieren, dass p* eine hinreichende, physische Ursache hat.[41] Nennen wir diese hinreichende physische Ursache p und gehen wir davon aus, dass p zeitgleich mit m stattfindet. Aufgrund der These der Nicht-Identität können wir davon ausgehen, dass das mentale Ereignis m nicht mit der hinreichenden Ursache p (oder mit einem Teil derselben) identisch ist. Wenn m also ebenfalls p* verursacht, dann hat p* nun zwei gleichzeitige Ursachen: Die hinreichende physische Ursache p und die zusätzliche mentale Ursache m. Diese Situation ist in Abbildung 3.3 illustriert.

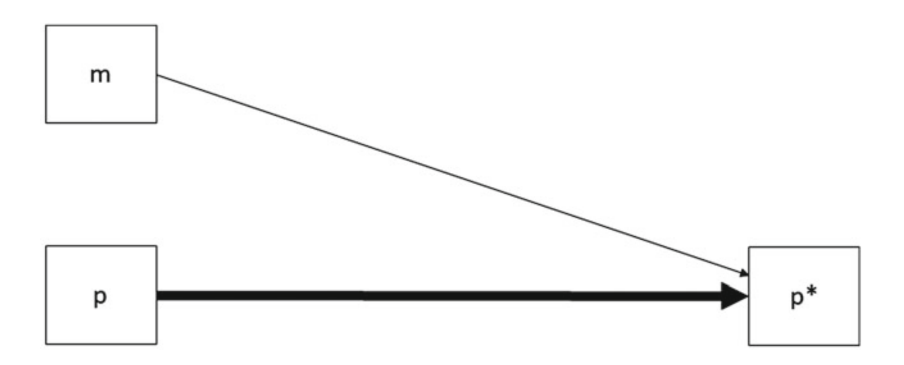
Diese Situation ergibt sich unmittelbar aus der These der Nicht-Identität und der These der kausalen Geschlossenheit. Zudem ergibt sich die Situation für jedes physische Ereignis, das laut der These der kausalen Geschlossenheit eine hinreichende physische Ursache hat. Jede Position, die – wie der nicht-reduktive Physikalismus – die These der Nicht-Identität und die These der kausalen

[38] Vgl. für genauere Formulierung und Diskussion meine Ausführungen in Abschnitt 2.4.2.

[39] Vgl. für genauere Formulierungen und Diskussion meine Ausführungen in Abschnitt 3.2.1.

[40] Die Bedenken, die ich in Abschnitt 3.2.5 hinsichtlich der Auffassung angemeldet habe, dass nicht-reduktive Physikalist*innen auf die These der kausalen Geschlossenheit festgelegt sind, können hier vorerst ignoriert werden.

[41] Tatsächlich wird hier die versteckte Annahme gemacht, dass p* überhaupt eine hinreichende Ursache hat und somit in den Anwendungsbereich der These der kausalen Geschlossenheit fällt. Anderenfalls würde die These der kausalen Geschlossenheit nicht motivieren können, dass p* eine hinreichende *physische* Ursache hat. Dies ist eines der zahlreichen Detailprobleme bei der Formulierung von Exklusionsargumenten, die ich hier übergehe (und die in den üblichen Formulierungen in der Literatur ebenfalls übergangen werden).



Erläuterung: Der dicke Pfeil steht für die Relation der hinreichenden Verursachung. Der dünnere Pfeil steht für die Relation der (hinreichenden) Verursachung.
p und p* sind physische$_{\text{eng}}$ Ereignisse. m ist ein mentales Ereignis.

Abbildung 3.3 Mental-physische Verursachung als Überdetermination

Geschlossenheit akzeptiert, ist also darauf festgelegt, dass mental-physische Verursachung, so es sie denn gibt, nach dem in Abbildung 3.3 abgebildeten Schema funktioniert.[42] Die physischen Wirkungen mentaler Ereignisse haben durchweg hinreichende physische Ursachen. Sie sind daher in einem gewissen Sinne überdeterminiert: Sie haben hinreichende Ursachen und zusätzlich weitere Ursachen.[43]

Bis hierher ist jedoch noch nicht ausgeschlossen, dass mentale Ereignisse überhaupt physische Wirkungen haben. Es ist lediglich ausgeschlossen, dass mentale Ereignisse physische Wirkungen haben, die keine hinreichenden physischen Ursachen haben. Es ist ja durchaus eine Möglichkeit, dass alle physischen Wirkungen mentaler Ereignisse in einem gewissen Sinne überdeterminiert sind.

Um also zu der Konklusion zu gelangen, dass mentale Ereignisse im nicht-reduktiven Physikalismus keine physischen Wirkungen haben, müssen weitere Annahmen hinzugezogen werden. Üblicherweise geschieht dies durch die

[42] Das einfache Exklusionsargument ist daher nicht spezifisch auf den nicht-reduktiven Physikalismus ausgerichtet, sondern betrifft auch dualistische Positionen.

[43] Ich bleibe hier vorerst relativ vage in Bezug auf den Begriff der Überdetermination. Weitere Diskussion und Präzisierung findet sich in 3.3.9.

Annahme eines allgemeinen Exklusionsprinzips, das hier vorläufig wie folgt formuliert werden kann:

> **Das einfache Exklusionsprinzip**: Wenn ein Ereignis eine hinreichende Ursache zu t hat, dann hat es darüber hinaus keine weitere Ursache zu t.[44]

Dieses einfache Exklusionsprinzip schließt nun aus, dass physische Ereignisse eine hinreichende physische Ursache und darüber hinaus eine (gleichzeitige) mentale Ursache haben. Wenn p* in Abbildung 3.3 also bereits eine hinreichende physische Ursache hat – was durch die These der kausalen Geschlossenheit vorgegeben ist – dann kann m keine Ursache für p* mehr sein. Mental-physische Verursachung kann dementsprechend nicht stattfinden, wenn die These der kausalen Geschlossenheit, die These der Nicht-Identität und das einfache Exklusionsprinzip wahr sind. Mentale Ereignisse haben dann keine physischen Wirkungen.

Man beachte, dass das einfache Exklusionsprinzip zunächst neutral gegenüber der Unterscheidung zwischen mentalen und physischen Ursachen ist. Das Exklusionsprinzip schreibt für sich genommen nicht vor, dass p die eigentliche (hinreichende) Ursache für p* ist. Es wäre ebenso mit dem Exklusionsprinzip vereinbar, an m als der eigentlichen Ursache festzuhalten und die Annahme, dass p eine hinreichende Ursache für p* ist, aufzugeben. Das Exklusionsprinzip verlangt lediglich eine Entscheidung zwischen den beiden Kandidaten für Ursachen: Eine der beiden Ursachen muss aussortiert werden. Ausschlaggebend dafür, dass m als Ursache aussortiert wird, ist die These der kausalen Geschlossenheit. Denn diese schreibt vor, dass p* eine hinreichende physische Ursache hat. Es ist also die These der kausalen Geschlossenheit, die eine Asymmetrie zwischen den beiden Kandidaten für Ursachen von p* ins Spiel bringt und eine Entscheidung zugunsten der physischen Ursache herbeiführt.[45]

---

[44] Das einfache Exklusionsprinzip ist absichtlich eine *vereinfachte* Form des Exklusionsprinzips, das etwa Kim (2005: 17) verwendet: „If an event e has a sufficient cause c at t, no event at t distinct from c can be a cause of e (unless this is a genuine case of causal overdetermination)". Ähnliche (vorläufige) Formulierungen des Exklusionsprinzips finden sich in Carey (2011: 252), Menzies (2013: 64) und Morris (2019: 63). Die Diskussion komplexerer und alternativer Exklusionsprinzipien, die insbesondere auch die in Kims Prinzip enthaltene Ausnahmeklausel für Fälle von genuiner Überdetermination berücksichtigen, verschiebe ich auf Abschnitt 3.3.9. Die Ausnahmeklausel macht einen zusätzlichen Schritt in den bis dahin diskutierten Exklusionsargumenten nötig, lässt ihre Struktur aber in allen anderen Hinsichten unberührt.

[45] Vgl. Kim (2005: 43).

Zur Zusammenfassung und Verdeutlichung der Argumentation möchte ich die folgende Rekonstruktion des einfachen Exklusionsarguments einführen:

Das einfache Exklusionsargument
Es sei m ein mentales Ereignis. p und p* seien zwei physische Ereignisse. Unter diesen Annahmen lässt sich das einfache Exklusionsargument wie folgt rekonstruieren:

(1) p ist hinreichende Ursache für p* (motiviert durch die These der kausalen Geschlossenheit).
(2) m ist nicht mit p (oder einem Teil von p) identisch (motiviert durch die These der Nicht-Identität).
(3) Wenn p hinreichende Ursache von p* ist und m nicht mit p (oder einem Teil von p) identisch ist, dann ist m nicht Ursache von p* (motiviert durch das einfache Exklusionsprinzip).
(4) Also: m ist nicht Ursache von p* (aus 1, 2, 3).

3.3.2 Die Reichweite des einfachen Exklusionsarguments

Erinnern wir uns an Susi und die vier Behauptungen mentaler Verursachung aus Abschnitt 3.1.1.:

Mental-mentale Verursachung: Susis interessanter Gedanke verursacht ihren Wunsch, die Aufmerksamkeit der Dozentin auf sich zu lenken.

Mental-behaviorale Verursachung: Susis Wunsch verursacht, dass sie ihren Arm hebt.

Mental-neurologische Verursachung: Susis Wunsch verursacht neurologische Veränderungen in ihrem motorischen Cortex.

Mental-mikrophysikalische Verursachung: Susis Wunsch verursacht mikrophysikalische Veränderungen in den Elementarteilchen, die ihren neurologischen Cortex konstituieren.

Die Frage nach der Reichweite des einfachen Exklusionsarguments, die ich in diesem Abschnitt behandeln möchte, ist diese: Welche dieser vier Behauptungen mentaler Verursachung sind durch das einfache Exklusionsargument betroffen?

Die kurze Antwort ist, dass von diesen vier Behauptungen ausschließlich die Behauptung mental-mikrophysikalischer Verursachung betroffen ist. Das einfache Exklusionsargument vermag nicht in Frage zu stellen, dass Susis Gedanke ihren Wunsch verursacht und es vermag nicht in Frage zu stellen, dass ihr Wunsch

ihre Armbewegung und neurologische Vorgänge in ihrem motorischen Cortex verursacht.

Der Grund dafür ist folgender: Das einfache Exklusionsargument beruft sich auf die These der kausalen Geschlossenheit. Es ist diese These, die erlaubt, die Existenz einer hinreichenden physischen Ursache von p* zu postulieren. Die Reichweite des einfachen Exklusionsarguments ist daher von der Reichweite der These der kausalen Geschlossenheit abhängig: Die These der kausalen Geschlossenheit erlaubt uns nur dann, eine hinreichende Ursache von p* zu postulieren, wenn p* überhaupt in den Bereich der These der kausalen Geschlossenheit fällt.

In Abschnitt 3.2.3 habe ich dafür argumentiert, dass die These der kausalen Geschlossenheit primär physische$_{\text{eng}}$ Ereignisse betrifft: Die Version der These der kausalen Geschlossenheit, die gut begründet ist und für die es plausibel scheint, dass nicht-reduktive Physikalist*innen auf sie festgelegt sind, besagt, dass *physische$_{eng}$* Ereignisse (die überhaupt hinreichende Ursachen haben) hinreichende, *physische$_{eng}$* Ursachen haben. Die These der kausalen Geschlossenheit kann also allenfalls dafür genutzt werden, hinreichende Ursachen für Ereignisse zu postulieren, die in den Bereich der Physik fallen. Weder neurologische noch behaviorale Ereignisse fallen aber in den Bereich der Physik. Diese Ereignisse sind zwar physisch$_{\text{weit}}$, aber eben nicht physisch$_{\text{eng}}$. Daher kann die These der kausalen Geschlossenheit auch nicht dafür genutzt werden, für diese Ereignisse hinreichende Ursachen zu postulieren.[46]

Dies ist eine bedeutende Limitierung des einfachen Exklusionsarguments. Denn nicht-reduktive Physikalist*innen könnten die Konklusion des Arguments einfach akzeptieren, da die im Alltagsdenken zentralen Behauptungen mentaler Verursachung noch gar nicht angegriffen sind. Um ein ernstzunehmendes Problem für den nicht-reduktiven Physikalismus darzustellen, sollte die Reichweite des einfachen Exklusionsarguments also ausgeweitet werden. Im Folgenden gehe ich auf einige Möglichkeiten ein, wie dies geschehen kann.

---

[46] Wer mir in diesem Punkt, für den ich im Abschnitt 3.2.3. argumentiert habe, nicht folgt, sollte noch immer den folgenden Punkt kaufen: Die These der kausalen Geschlossenheit ist zumindest nicht ohne weitere Argumentation uneingeschränkt allgemein. Sie erlaubt daher nicht, für *alle* Ereignisse hinreichende physische Ursachen zu postulieren. Jede Einschränkung der Reichweite der These der kausalen Geschlossenheit resultiert aber in einer Einschränkung der Reichweite des einfachen Exklusionsarguments. Das einfache Exklusionsargument ist also nicht uneingeschränkt allgemein und muss ausgebaut werden, um zu etablieren, dass mentale Ereignisse keinerlei Wirkungen haben.

3.3.3 Kausale Geschlossenheit und empirische Evidenzen für hinreichende physische Ursachen

Die Einschränkung der Reichweite des einfachen Exklusionsarguments ergibt sich aus der eingeschränkten Reichweite der These der kausalen Geschlossenheit. Ein erster Ansatz, diese Einschränkung aufzulösen, besteht also darin, das einfache Exklusionsargument von der These der kausalen Geschlossenheit zu lösen.

Tatsächlich ist die These der kausalen Geschlossenheit in einer Hinsicht sehr viel allgemeiner, als dies im Rahmen des einfachen Exklusionsargument verlangt wird. Denn die These der kausalen Geschlossenheit betrifft *alle* $\text{physischen}_{\text{eng}}$ Ereignisse und nicht nur jene $\text{physischen}_{\text{eng}}$ Ereignisse, die als Wirkungen mentaler Ereignisse in Frage kommen. Die These der kausalen Geschlossenheit wäre zum Beispiel falsch, wenn die physische Welt einen nicht-physischen kausalen Ursprung hätte: Wenn ein übernatürliches Wesen den Urknall (als erstes physisches Ereignis) ausgelöst hätte, gäbe es ein physisches Ereignis, das eine hinreichende Ursache hat, aber keine hinreichende physische Ursache. Das einfache Exklusionsargument würde dann scheitern – und zwar aufgrund von Erwägungen, die mit der Existenz gewöhnlicher mentaler Verursachung eigentlich nichts zu tun haben.[47] Es ist daher wünschenswert, das einfache Exklusionsargument von der unnötigen Allgemeinheit der These der kausalen Geschlossenheit zu befreien.

Anstatt sich also auf ein allgemeines Prinzip zu berufen, dass es erlaubt, eine hinreichende physische Ursache von p* zu postulieren, könnte man sich auf konkrete empirische Belege dafür berufen, dass p* eine hinreichende physische Ursache hat. Man könnte Prämisse (1) im einfachen Exklusionsargument („p ist hinreichende Ursache für p*") also auch motivieren, indem man sich auf konkrete empirische Evidenzen aus der Physiologie oder den Neurowissenschaften beruft, anstatt sich auf eine allgemeine These der kausalen Geschlossenheit zu berufen: Wenn beispielsweise eine vollständige neuro-physiologische Kausalerklärung von Susis Armbewegung vorliegt, die die kausale Geschichte der Armbewegung über Muskelkontraktionen, Aktivitäten im motorischen Cortex und darüber hinaus lückenlos zurückverfolgt, motiviert bereits dies auf den ersten Blick die Annahme, dass die Armbewegung eine hinreichende (neurologische) Ursache hat. Für die Plausibilität dieser Annahme ist es dann gar nicht notwendig, eine allgemeine These der kausalen Geschlossenheit zu verteidigen, die neben behavioralen Ereignissen auch alle anderen physischen Ereignisse betrifft.

[47] Vgl. Hüttemann (2013b: 43).

Eine solche Herangehensweise hat (möglicherweise) wünschenswerte Konsequenzen für die Reichweite des Arguments: Betroffen sind nun alle Ereignisse, für die es empirisch plausibel ist, dass sie hinreichende $\text{physische}_{\text{weit}}$ Ursachen haben. Eine prinzipielle Einschränkung auf $\textit{physische}_{\textit{eng}}$ Ereignisse liegt nicht mehr vor.

Ob die Reichweite des einfachen Exklusionsarguments durch diese Alternative tatsächlich auf $\text{physische}_{\text{weit}}$ Ereignisse, insbesondere behaviorale Ereignisse, ausgeweitet werden kann, hängt natürlich davon ab, ob solche vollständigen Kausalerklärungen der relevanten $\text{physischen}_{\text{weit}}$ Ereignisse durch andere $\text{physische}_{\text{weit}}$ Ereignisse verfügbar sind. Diese Frage werde ich hier nicht im Detail behandeln. Auf drei mögliche Schwierigkeiten dieser Strategie sei jedoch hingewiesen:

Erstens: Was im Rahmen des Exklusionsarguments benötigt wird, ist die Annahme, dass die relevanten $\text{physischen}_{\text{weit}}$ Ereignisse *hinreichende* $\text{physische}_{\text{weit}}$ Ursachen haben. Es würde also nicht reichen, $\text{physische}_{\text{weit}}$ *Teilursachen* ausfindig zu machen. Denn auch die These der kausalen Exklusion wird letztlich bestenfalls dann plausibel, wenn sie sich auf hinreichende Verursachung bezieht.[48] Die These, dass behaviorale Ereignisse hinreichende $\text{physische}_{\text{weit}}$ Ursachen haben, ist jedoch deutlich schwieriger empirisch zu belegen, als die schwächere These, dass behaviorale Ereignisse $\text{physische}_{\text{weit}}$ Ursachen haben.[49] Zumindest würde die Identifikation einer relativ lokalen neurologischen Ursache hierfür nicht ausreichen. Denn angesichts der allgemeinen Komplexität hinreichender Ursachen ist eine solche relativ lokale neurologische Ursache kein guter Kandidat für eine hinreichende Ursache.[50]

Zweitens: Die These der kausalen Geschlossenheit übernimmt im einfachen Exklusionsargument die Rolle, eine Entscheidung zwischen zwei möglichen Ursachen herbeizuführen. Denn das Exklusionsprinzip schreibt lediglich vor, dass entweder m oder p keine Ursache von p* sein kann. Die These der kausalen Geschlossenheit schließt dann die Option aus, p als Ursache aufzugeben und an m festzuhalten.[51] Da die These der kausalen Geschlossenheit im Rahmen von

---

[48] Vgl. auch meine Diskussion in Abschnitt 4.2.3 für eine Begründung dieser Behauptung.

[49] Vgl. auch Hüttemann (2013b: fn5).

[50] Natürlich könnte man sich hier wieder auf die Behauptung berufen, dass lediglich in den Umständen hinreichende Ursachen benötigt werden. Diese Strategie hat jedoch das grundsätzliche Problem, dass die These der kausalen Exklusion letztlich auch unter Voraussetzung eines Verständnisses von hinreichenden Ursachen als in den Umständen hinreichenden Ursachen nicht plausibel wird.

[51] Vgl. Kim (2005: 43): „Exclusion itself is neutral with respect to the mental-physical competition; it says either the mental cause or the physical cause must go, but doesn't favor either

Diskussionen des Exklusionsarguments gegen den nicht-reduktiven Physikalismus als unverhandelbar gilt, wird die Option, an dieser Stelle des Arguments einzuhaken, üblicherweise nicht wahrgenommen. Es stellt sich jedoch die Frage, ob die Berufung auf empirische Evidenzen dieselbe Funktion innerhalb von einem Exklusionsargument gegen den nicht-reduktiven Physikalismus übernehmen kann wie die These der kausalen Geschlossenheit. Denn es ließe sich argumentieren, dass die empirischen Evidenzen für die Existenz mentaler Verursachung mindestens ebenso stark sind wie die empirischen Evidenzen für die Existenz hinreichender physischer Ursachen. Zunächst einmal etabliert die These der kausalen Exklusion dann einen Konflikt zwischen empirischen Evidenzen unterschiedlicher Art. In diesem Fall ist es gar nicht so klar, dass nicht-reduktive Physikalist*innen die Evidenzen für die Existenz hinreichender physischer Ursachen bevorzugen müssen. Es muss im Rahmen dieser Strategie also argumentiert werden, dass die empirischen Evidenzen für die Existenz hinreichender physischer Ursachen für die vermeintlichen Wirkungen mentaler Ereignisse nicht nur *vorliegen*, sondern sie die empirischen Evidenzen für die Existenz mentaler Verursachung auch *übertrumpfen*.

Drittens: Die These der kausalen Geschlossenheit erlaubt es, für eine große Klasse von Ereignissen auf einen Schlag hinreichende Ursachen zu postulieren: Alle physischen$_{eng}$ Ereignisse sind daher gleichermaßen vom einfachen Exklusionsargument betroffen. Es ist nicht nötig, auf die Details der kausalen Geschichte bestimmter physischer$_{eng}$ Ereignisse einzugehen, um hinreichende Ursachen für sie zu postulieren. Exklusionsargumente, die auf konkreten empirischen Evidenzen für hinreichende Ursachen der vermeintlichen Wirkungen mentaler Ereignisse beruhen, müssen hingegen viel stärker stückchenweise vorgehen: Ein Exklusionsargument, das etabliert, das Susis Gedanke nicht ihren Wunsch verursacht, muss auf empirische Theorien zur kausalen Geschichte von Wünschen eingehen. Ein Exklusionsargument, das etabliert, dass Susis Wunsch nicht ihre Armbewegung verursacht, muss auf empirische Theorien zur kausalen Geschichte von Armbewegungen eingehen. Und ein Exklusionsargument, das etabliert, dass Susis Wunsch keine neurologischen Veränderungen in ihrem motorischen Kortex verursacht, muss auf empirische Theorien über die kausale Geschichte entsprechender Vorgänge im motorischen Kortex eingehen. Um eine universelle epiphänomenalistische Konklusion zu erhalten, müssten daher für jeden Typ von vermeintlichen Wirkungen mentaler Ereignisse gut bestätige Theorien vorliegen. Es scheint daher

over the other. What makes the difference – what introduces an asymmetry into the situation – is Closure. It is the causal closure of the physical world that excludes the mental cause, enabling the physical cause to prevail.“

zumindest einfacher zu sein, allgemeine Exklusionsargumente auf der Verteidigung eines allgemeinen Prinzips wie der These der kausalen Geschlossenheit aufzubauen.

Im Lichte dieser (sicherlich nicht vernichtenden) Schwierigkeiten lohnt es sich also, nach weiteren Möglichkeiten Ausschau zu halten, die Reichweite des einfachen Exklusionsarguments zu erweitern.

3.3.4 Mentale Verursachung im Stufenmodell der Realität

Das einfache Exklusionsargument beruht auf der These der Nicht-Identität und der These der kausalen Geschlossenheit. Es enthält keine darüber hinausgehenden, für den nicht-reduktiven Physikalismus spezifischen Thesen. Daher betrifft es neben dem nicht-reduktiven Physikalismus auch dualistische Positionen. Die in den folgenden Abschnitten zu diskutierenden Erweiterungen weichen in dieser Hinsicht vom einfachen Exklusionsargument ab. Denn sie berufen sich auch auf die für den nicht-reduktiven Physikalismus spezifische These der ontologischen Abhängigkeit:

> **Universale Ontologische Abhängigkeit**: Jedes höherstufige Ereignis h ist ontologisch abhängig von einem physischen_{eng} Ereignis p.[52]

Um die verschiedenen Arten, wie diese These für Exklusionsargumente relevant wird, zu explizieren, ist es an dieser Stelle sinnvoll, ein wenig Terminologie einzuführen und die Existenz mentaler Verursachung auf diese Weise etwas genauer innerhalb der Metaphysik des nicht-reduktiven Physikalismus zu verorten.

Wie in Abschnitt 2.5.8. ausgeführt, zieht der nicht-reduktive Physikalismus eine erschöpfende und ausschließende Unterscheidung zwischen fundamentalen, physischen_{eng} Ereignissen auf der einen Seite und höherstufigen Ereignissen auf der anderen Seite. Höherstufige Ereignisse sind dabei ontologisch abhängig von fundamentalen Ereignissen. In Abschnitt 2.5.9 habe ich zudem die Idee eines Stufenmodells der Realität ausgeführt: Höherstufige Ereignisse sind diesem Modell zufolge hierarchisch in niedrigere und höhere Stufen geordnet. Die verschiedenen Stufen lassen sich grob den unterschiedlichen Spezialwissenschaften zuordnen, so dass es eine (mikro-)physikalische Stufe, eine chemische Stufe, eine biologische Stufe, eine neurologische Stufe, eine psychologische (oder mentale) Stufe

[52] Vgl. für genauere Formulierungen und Diskussionen meine Ausführungen in Abschnitt 2.4.

gibt usw.. Die Hierarchie ergibt sich dabei aus der Relation der ontologischen Abhängigkeit: Die Idee ist also, dass mentale Ereignisse auf einer höheren Stufe als neurologische Ereignisse sind, weil mentale Ereignisse von neurologischen Ereignissen ontologisch abhängen. Neurologische Ereignisse sind hingegen auf einer höheren Stufe als biologische Ereignisse, weil neurologische Ereignisse von biologischen Ereignissen ontologisch abhängen. Auf der absoluten Basis dieses Modells[53] finden wir fundamentale, physische$_{eng}$ Ereignisse. Alle höherstufigen Ereignisse hängen letztlich von diesen physischen$_{eng}$ Ereignisse ontologisch ab.

Vor dem Hintergrund eines Stufenmodells der Realität lässt sich nun eine Unterscheidung zwischen drei Arten von Verursachung treffen:

> *Intralevel Verursachung* liegt genau dann vor, wenn Ursache und Wirkung sich auf derselben Stufe befinden – wenn also zum Beispiel sowohl Ursache als auch Wirkung neurologische Ereignisse sind.
>
> *Abwärtsverursachung* liegt genau dann vor, wenn die Ursache sich auf einer höheren Stufe befindet als die Wirkung – wenn also zum Beispiel die Ursache ein mentales Ereignis und die Wirkung ein neurologisches Ereignis ist.
>
> *Aufwärtsverursachung* liegt genau dann vor, wenn die Ursache sich auf einer niedrigeren Stufe befindet als die Wirkung – wenn also zum Beispiel die Ursache ein neurologisches Ereignis und die Wirkung ein mentales Ereignis ist.

In Abbildung 3.4 ist diese Unterscheidung anhand einer niedrigeren und einer höheren Stufe illustriert.

Ein wichtiger Spezialfall von Intralevel-Verursachung ist dabei die *fundamentale Verursachung* (siehe Abb. 4.2), bei der ein fundamentales, physisches$_{eng}$ Ereignis ein anderes fundamentales, physisches$_{eng}$ Ereignis verursacht. Es ist diese Art der Verursachung, die primär durch die These der kausalen Geschlossenheit gedeckt ist: Fundamentale, physische$_{eng}$ Ereignisse haben (hinreichende) fundamentale, physische$_{eng}$ Ursachen.[54]

Wie ist nun mentale Verursachung in diese Unterscheidung eingebettet? Wie in Abschnitt 3.1.1 gesehen, gilt es zunächst zu beachten, dass es ganz unterschiedliche Behauptungen mentaler Verursachung gibt. In der Geschichte von Susi, die einen interessanten Gedanken hat und deshalb ihren Arm hebt, um die Aufmerksamkeit der Dozentin auf sich zu lenken, habe ich vier mögliche Behauptungen mentaler Verursachung unterschieden:

---

[53] Ich mache hier – wie in Abschnitt 2.5.8. auch schon erwähnt – die vereinfachende Annahme, dass es eine solche absolut fundamentale Basis gibt.

[54] Siehe Abschnitt 3.2.3.

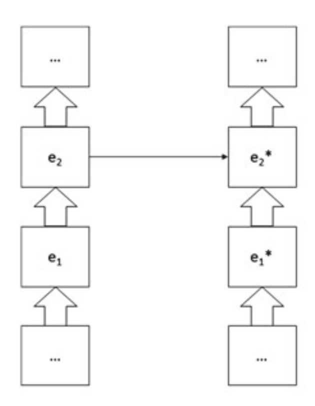


Abb. 4.1. Intralevel-Verursachung

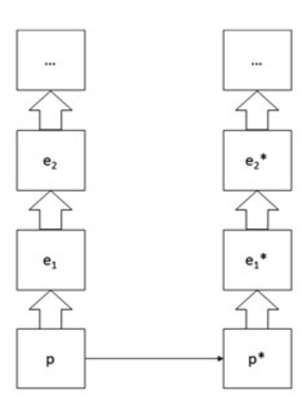


Abb. 4.2. Fundamentale Verursachung als Spezialfall von Intralevel-Verursachung

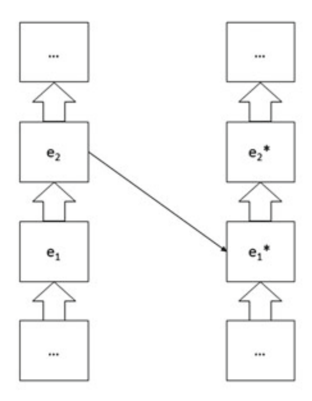


Abb. 4.3. Abwärtsverursachung

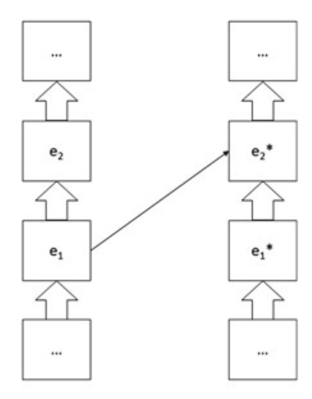


Abb. 4.4. Aufwärtsverursachung

Erläuterung: Die dicken Pfeile stehen für die Relation der ontologischen Abhängigkeit. Die dünnen Pfeile stehen für die Relation der (hinreichenden) Verursachung. Ereignisse e1 und e1* sind auf derselben Stufe, Ereignisse e2 und e2* sind auf derselben Stufe usw. Die Pünktchen (...) ober- und unterhalb machen deutlich, dass es zudem weitere, niedrigere und höhere Stufen gibt.

Abbildung 3.4 Verursachung im Stufenmodell der Realität

Mental-mentale Verursachung: Susis Gedanke verursacht ihren Wunsch.

Mental-behaviorale Verursachung: Susis Wunsch verursacht, dass sie ihren Arm hebt.

Mental-neurologische Verursachung: Susis Wunsch verursacht neurologische Veränderungen in ihrem motorischen Cortex.

Mental-mikrophysikalische Verursachung: Susis Wunsch verursacht mikrophysikalische Veränderungen in den Elementarteilchen, die ihren neurologischen Cortex konstituieren.

Diese verschiedenen Fälle mentaler Verursachung sind unterschiedlich in die in diesem Abschnitt eingeführte Unterscheidung zwischen Intralevel-, Abwärts- und Aufwärtsverursachung einzuordnen. Mental-mentale Verursachung etwa ist eindeutig als ein Fall von Intralevel-Verursachung zu klassifizieren: Susis Gedanke und ihr Wunsch sind auf derselben Stufe angesiedelt. Mental-neurologische Verursachung kann als Spezialfall von Abwärtsverursachung eingeordnet werden: Mentale Ereignisse sind ontologisch abhängig von neurologischen Ereignissen und daher auf einer höheren Stufe angesiedelt. Wenn ein mentales Ereignis also ein neurologisches Ereignis verursacht, befindet sich die Ursache auf einer höheren Stufe als die Wirkung. Ebenso ist mental-mikrophysikalische Verursachung ein Spezialfall von Abwärtsverursachung. Hierbei ist eine Besonderheit, dass das mentale Ereignis ein physisches$_{eng}$, fundamentales Ereignis verursacht. Es handelt sich also um einen Fall von *mental-fundamentaler Abwärtsverursachung*. Es ist dieser Spezialfall von Abwärtsverursachung, der im einfachen Exklusionsargument problematisiert wird.

Die Einordnung des im Alltagsdenken vielleicht zentralsten Falls der mental-behavioralen Verursachung ist hingegen weniger eindeutig. Einige Autor*innen würden dies als einen weiteren Spezialfall von Abwärtsverursachung einordnen: Behaviorale Ereignisse sind (im weiten Sinne) physische Ereignisse und schon deshalb auf einer niedrigeren Stufe als mentale Ereignisse. Andere verstehen mental-behaviorale Verursachung hingegen als einen Spezialfall von Intralevel-Verursachung. Behaviorale Ereignisse sind, genau wie mentale Ereignisse, höherstufig und von physischen$_{eng}$ Ereignissen ontologisch abhängig. Zudem ließe sich argumentieren, dass mentale und behaviorale Ereignisse derselben Spezialwissenschaft, nämlich der Psychologie, zugeordnet sind und sie deshalb als gleichstufig behandelt werden sollten.[55] Hier hängt die Einordung also von kontroversen Fragen über den konkreten Aufbau der Stufenhierarchie ab. Wie wir sehen werden, hat die Einordnung mental-behavioraler Verursachung

[55] Vgl. Gibbons (2006).

bedeutende Konsequenzen für die Plausibilität mancher Erwiderungen auf das Exklusionsargument.[56]

Es bleibt also zunächst festzuhalten, dass die ausschlaggebenden Behauptungen mentaler Verursachung entweder Fälle von mentaler Intralevel-Verursachung oder Fälle von mentaler Abwärtsverursachung sind.[57] Mental-fundamentale Abwärtsverursachung nimmt dabei einen Sonderstatus ein, da es eben jene Art von Verursachung ist, deren Existenz durch das einfache Exklusionsargument in Frage gestellt wird. Die noch zu diskutierenden Erweiterungen des einfachen Exklusionsarguments nehmen die Problematisierung von mental-fundamentaler Abwärtsverursachung als Ausgangspunkt, um auch andere Arten mentaler Verursachung zu problematisieren. Dabei berufen sie sich auf eines von drei Prinzipien der Verursachung im Stufenmodell der Realität. Diese drei Prinzipien werden im folgenden Abschnitt erläutert.

3.3.5 Drei Prinzipien der Verursachung im Stufenmodell der Realität

Im Stufenmodell der Realität treffen zwei unterschiedliche Abhängigkeits-Relationen aufeinander: Auf der einen Seite haben wir die ‚vertikale' Relation der ontologischen Abhängigkeit, die gleichzeitig stattfindende, fundamentale und höherstufige Ereignisse miteinander verbindet. Auf der anderen Seite haben wir die ‚horizontale' Relation der Verursachung, die (üblicherweise) zeitlich aufeinanderfolgende Ursache- und Wirkungsereignisse verknüpft. Es stellt sich nun die Frage, wie diese beiden Relationen miteinander interagieren. Im vorliegenden Kontext ist besonders die folgende Frage relevant: Was folgt daraus, dass ein Ereignis von einem anderen Ereignis ontologisch abhängt, für das Bestehen oder Nicht-Bestehen von Verursachungsrelationen, die diese Ereignisse involvieren?

Drei Prinzipien, die diese Frage teilweise beantworten, spielen für die zu diskutierenden Erweiterungen des einfachen Exklusionsarguments eine wichtige

[56] Betroffen sind insbesondere Erwiderungen, die die Existenz von mentaler Intralevel-Verursachung verteidigen, aber die Existenz von mentaler Abwärtsverursachung ablehnen. Vgl. z. B. Gibbons (2006: 97 ff.), Zhong (2014: 350) und Pernu (2016: 1048) für entsprechende Anmerkungen.

[57] Der Vollständigkeit halber sollte man auch die Möglichkeit mentaler Aufwärtsverursachung erwähnen: Diese könnte etwa vorliegen, wenn mentale Ereignisse ökonomische Ereignisse verursachen: Eine Entscheidung verursacht die Anhebung des Leitzinses.

Rolle: Das Prinzip der Abwärtsverursachung, das Prinzip der Aufwärtsverursachung und das Prinzip der basalen Verursachung. Diese drei Prinzipien werde ich im Folgenden jeweils formulieren, erläutern und knapp und vorläufig motivieren.

Das *Prinzip der Abwärtsverursachung* besagt das folgende:

> **Prinzip der Abwärtsverursachung**: Wenn ein Ereignis u eine (hinreichende)[58] Ursache für ein Ereignis w ist und Ereignis w ontologisch abhängig von einem Ereignis b ist, dann ist u auch eine (hinreichende) Ursache für Ereignis b.[59]

Jede Ursache eines höherstufigen Ereignisses ist diesem Prinzip zufolge zugleich Ursache aller niedrigerstufigen Ereignisse, von denen das höherstufige Ereignis ontologisch abhängt. Die Relation der Verursachung sickert in der Hierarchie der Stufen gewissermaßen herunter.

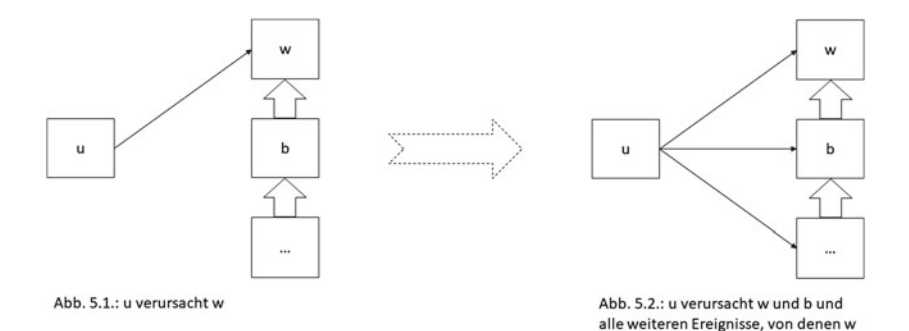


Abbildung 3.5 Prinzip der Abwärtsverursachung

[58] Ich formuliere die drei Prinzipien in diesem Abschnitt neutral hinsichtlich der Frage, ob sie die Relation der *hinreichenden* Verursachung oder die Relation der *Verursachung* betreffen. Diese Frage ist jedoch, wie noch deutlich werden wird, sowohl relevant für die Plausibilität der Prinzipien als auch für die Stärke der aus ihnen resultierenden Erweiterungen des einfachen Exklusionsarguments. Ich werde daher in Abschnitt 4.2.4. noch einmal darauf eingehen.

[59] Vgl. z. B. Kim (1993b: 205–206): „If a given instance of S occurs by being realized by Q, then any cause of this instance of S must be a cause of this instance of Q“; Zhong (2011: 138): „*Downward Causation*: If property A causes property B, then A must cause any base property of B instantiated on this occasion“.

In Abbildung 3.5 ist illustriert, welche Art von Schlüssen das Prinzip der Abwärtsverursachung erlaubt. Die Situation in Abbildung 5.1. impliziert nach dem Prinzip der Abwärtsverursachung die Situation in Abbildung 5.2.. Abbildung 5.1. kann daher niemals eine vollständige kausale Struktur abbilden. Stattdessen ist die vollständige Struktur nur durch Abbildung 5.2. abgebildet.

Eine wichtige Konsequenz des Prinzips der Abwärtsverursachung ist, dass es eine *reine* Intralevel-Verursachung nur auf der fundamentalen Stufe geben kann. In anderen Worten: Intralevel-Verursachung geht auf höheren Stufen immer mit Abwärtsverursachung einher. Wenn Susis Gedanke etwa ihren Wusch verursacht (mental-mentale Intralevel-Verursachung), dann muss ihr Gedanke auch das neurologische Ereignis verursachen, von dem ihr Wunsch ontologisch abhängt (mental-neurologische Abwärtsverursachung). Aber nicht nur Intralevel-Verursachung ist vom Prinzip der Abwärtsverursachung betroffen. Eine weitere wichtige Konsequenz ist, dass Abwärtsverursachung immer herunter bis auf die fundamentale Stufe sickert. Wenn Susis Wunsch also neurologische Vorgänge in ihrem motorischen Cortex verursacht (mental-neurologische Abwärtsverursachung), dann muss ihr Wunsch auch die mikrophysikalischen Veränderungen in den Elementarteilchen verursachen, die ihren motorischen Cortex konstituieren (mental-fundamentale Abwärtsverursachung).

Kim verteidigt dieses Prinzip im Rahmen seiner bevorzugten Erweiterung des einfachen Exklusionsarguments. Er stellt hierfür zunächst eine Spannung fest, die sich aus zwei verschiedenen Erklärungen des Eintretens von w ergibt: Die ‚horizontale' Kausalerklärung von w, die auf die Ursache u verweist, steht in Spannung mit der ‚vertikalen' Erklärung von w, die auf das niedrigerstufige Ereignis b verweist. Diese Spannung könne durch die Annahme aufgelöst werden, dass u auch b verursacht. Dies ist also eine vorläufige Motivation für das Prinzip der Abwärtsverursachung. Sie ergibt sich aus einer behaupteten Spannung zwischen den beiden Abhängigkeits-Relationen, die im Stufenmodell der Realität aufeinandertreffen: der horizontalen Relation der Verursachung auf der einen Seite und der vertikalen Relation der ontologischen Abhängigkeit auf der anderen Seite.[60]

Das Prinzip der Aufwärtsverursachung kann wie folgt formuliert werden:

[60] Kim (2005: 39–40). Vgl. auch Kim (1993: 203–206).

Prinzip der Aufwärtsverursachung: Wenn ein Ereignis u eine (hinreichende) Ursache für ein Ereignis w ist und ein Ereignis h ontologisch abhängig von Ereignis w ist, dann ist u auch eine (hinreichende) Ursache für Ereignis h.[61]

Jede Ursache eines niedrigerstufigen Ereignisses ist diesem Prinzip zufolge zugleich Ursache aller höherstufigen Ereignisse, die von dem niedrigerstufigen Ereignis ontologisch abhängen. Die Relation der Verursachung klettert in der Hierarchie der Stufen gewissermaßen hinauf.

In Abbildung 3.6 ist illustriert, welche Arten von Schlüssen das Prinzip der Aufwärtsverursachung erlaubt. Die Situation in Abbildung 6.1. impliziert nach dem Prinzip der Aufwärtsverursachung die Situation in Abbildung 6.2.. Abbildung 6.1. kann daher niemals eine vollständige kausale Struktur abbilden. Stattdessen ist die vollständige Struktur nur durch Abbildung 6.2. abgebildet.

Ebenso wie das Prinzip der Abwärtsverursachung hat auch das Prinzip der Aufwärtsverursachung Konsequenzen für die Existenz reiner Intralevel-Verursachung. Wenn das Prinzip der Aufwärtsverursachung stimmt, kann es eine reine Intralevel-Verursachung nur auf einer höchsten Stufe der Realität geben: Nur, wenn von der Wirkung keine weiteren Ereignisse ontologisch abhängen, folgt aus einer Intralevel-Verursachungsrelation keine Relation der Aufwärtsverursachung. Intralevel-Verursachung geht auf niedrigeren Stufen immer mit Aufwärtsverursachung einher. Wenn also die Veränderungen in den Elementarteilchen, die Susis motorischen Cortex konstituieren, von vorhergehenden mikrophysikalischen Vorgängen verursacht werden (fundamentale Intralevel-Verursachung), dann müssen diese mikrophysikalischen Vorgänge

[61] Das Prinzip wird häufiger verwendet, als es explizit benannt wird. Zwei Beispiele für Anwendungen dieses Prinzips in der Literatur sind die Folgenden: Erstens: „Indem p_1 allerdings p_2 verursacht, bewirkt p_1 auch das, was auf p_2 superveniert, nämlich m_2“ (Esfeld (2005: 51)). Zweitens: „Since M* supervenes on P*, the complete physical cause of P* is also a cause of M*“ (Baker (2009: 7)). Eine explizite Formulierung findet sich bei Zhong (2011: 135): „*Upward Causation*: If property A causes property B, then A would cause any supervenient property of B instantiated on this occasion.“ Siehe auch Kim (1993: 205–206), dessen ‚Causal Realization Principle‘ das Aufwärtskausalitätsprinzip und das Abwärtskausalitätsprinzip impliziert: „If a given instance of S occurs by being realized by Q, then any cause of this instance of S must be a cause of this instance of Q (*and of course any cause of this instance of Q is a cause of this instance of S*)“ (meine Hervorhebung). Kim geht auf den zweiten, von mir kursiv hervorgehobenen, Teil dieses Prinzips, der dem Aufwärtskausalitätsprinzip entspricht und den er für selbstverständlich hält, leider nicht detailliert ein. Auch spielt dieser Teil keine Rolle für seine Argumentation in diesem Aufsatz. Obgleich Kim selbst meines Wissens nirgendwo ein Exklusionsargument auf dem Prinzip der Aufwärtsverursachung aufbaut, scheinen sowohl Esfeld als auch Baker ihre Argumente aus dem Prinzip der Aufwärtsverursachung Kim zuzuschreiben.

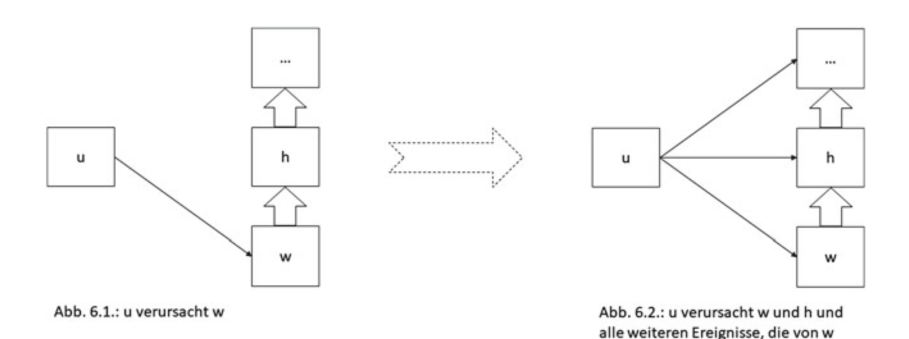


Abbildung 3.6 Prinzip der Aufwärtsverursachung

auch die neurologischen Vorgänge in Susis motorischem Cortex verursachen (fundamental-neurologische Aufwärtsverursachung). Auch das Prinzip der Aufwärtsverursachung hat zudem Konsequenzen für Abwärtsverursachung. Abwärtsverursachung klettert die Hierarchie der Stufen hinauf: Wenn Susis Wunsch die mikrophysikalischen Veränderungen in den Elementarteilchen verursacht, die ihren motorischen Cortex konstituieren (mental-fundamentale Abwärtsverursachung), dann muss ihr Wunsch auch die hierdurch konstituierten neurologischen Vorgänge in ihrem motorischen Cortex verursachen (mental-neurologische Abwärtsverursachung). In vielen Fällen bedeutet dies zudem, dass Abwärtsverursachung mit Intralevel-Verursachung einhergeht: Wenn Susis Gedanke etwa die neurologische Basis ihres Wunsches verursacht (mental-neurologische Abwärtsverursachung), dann verursacht ihr Gedanke auch ihren Wunsch (mental-mentale Intralevel-Verursachung).

Das Prinzip der Aufwärtsverursachung kann vorläufig durch die folgende Überlegung motiviert werden: Wenn u eine hinreichende Ursache für w ist, dann garantiert u, dass w eintritt. Wenn w aber eine physische Basis für h ist (weil h von w ontologisch abhängt), dann garantiert w auch, dass h eintritt. Also garantiert u ebenfalls, dass h eintritt. Dies spricht dafür, dass u auch h verursacht.[62]

[62] Diese Überlegung beruht natürlich einerseits darauf, dass u eine *hinreichende* Ursache für w ist und nicht bloß eine Teilursache. Zudem setzt diese Überlegung voraus, dass w eine

Das Prinzip der Aufwärtsverursachung und das Prinzip der Abwärtsverursachung haben zusammengenommen die Konsequenz, dass zwei Ereignisse, zwischen denen eine ontologische Abhängigkeitsbeziehung besteht, all ihre Ursachen teilen. Die Relation der Verursachung weitet sich dann auf das gesamte Spektrum der verschiedenen Stufen aus: Sie sickert nach unten und klettert nach oben. Verursachung bleibt nie ausschließlich auf einer Stufe. Wenn Susis Gedanke also ein neurologisches Ereignis verursacht, das als Basis für ihren Wunsch fungiert, dann verursacht der Gedanke sowohl ihren Wunsch (aufgrund des Prinzips der Aufwärtsverursachung) als auch die mikrophysikalischen Vorgänge, die das neurologische Ereignis konstituieren (aufgrund des Prinzips der Abwärtsverursachung).

Die beiden bis hierher besprochenen Prinzipien haben gemeinsam, dass sie von *Wirkungen* ausgehen, die in der Relation der ontologischen Abhängigkeit stehen. Sie haben hingegen keine Konsequenzen für Situationen, in denen *Ursachen* in der Relation der ontologischen Abhängigkeit stehen. Das dritte Prinzip, das für Erweiterungen des einfachen Exklusionsarguments genutzt werden kann und das es im Folgenden einzuführen gilt, geht nun von *Ursachen* aus, die in der Relation der ontologischen Abhängigkeit stehen.

Wir können dieses Prinzip ‚das Prinzip der basalen Verursachung' nennen und es wie folgt formulieren:

> **Prinzip der basalen Verursachung**: Wenn ein Ereignis u eine (hinreichende) Ursache für ein Ereignis w ist und Ereignis u ontologisch abhängig von einem Ereignis b ist, dann ist auch Ereignis b eine (hinreichende) Ursache für Ereignis w.[63]

---

vollständige Basis (ein ‚total realizer' in Shoemakers (1981) Terminologie) für h ist. Schließlich überzeugt diese Argumentation in erster Linie im Rahmen einer nomologischen (oder regularitätstheoretischen) Auffassung von Verursachung. Im Rahmen von konkreten Theorien der Verursachung lässt sich die Frage, ob sich das Prinzip auch ohne diese Annahmen plausibilisieren lässt, besser behandeln. Ich gehe in Teil 4 noch einmal auf diese Frage ein.

[63] Vgl. z. B. Kim (2005: 41), Kim (2011: 127–128) und Morris (2014: 74–77) für Anwendungen dieses Prinzips. Eine explizite Formulierung findet sich in Menzies (2013: 63): „*The transmission of causal sufficiency across realization*: if a mental state M is causally sufficient for a behavioural state B and M is realized by a distinct physical state P, then the physical state is causally sufficient for the behavioural state B". Verwandt ist auch Kims (1993: 208) ‚Principle of Causal Inheritance': „If M is instantiated on a given occasion by being realized by P, then the causal powers of this instance of M are identical with (perhaps, a subset of) the causal powers of P." Vgl. zudem Zhong (2011: 131–135) und Menzies (2013) für Exklusionsargumente, die sich auf ein solches Prinzip beziehen.

Jede Wirkung eines höherstufigen Ereignisses ist diesem Prinzip zufolge zugleich Wirkung jedes niedrigerstufigen Ereignisses, von dem das höherstufige Ereignis ontologisch abhängt. Wir haben hier also einen weiteren Sinn, in dem die Relation der Verursachung in der Hierarchie der Stufen heruntersickert: Was von einem höherstufigen Ereignis verursacht wird, wird auch von dessen niederigerstufigen Basen verursacht.

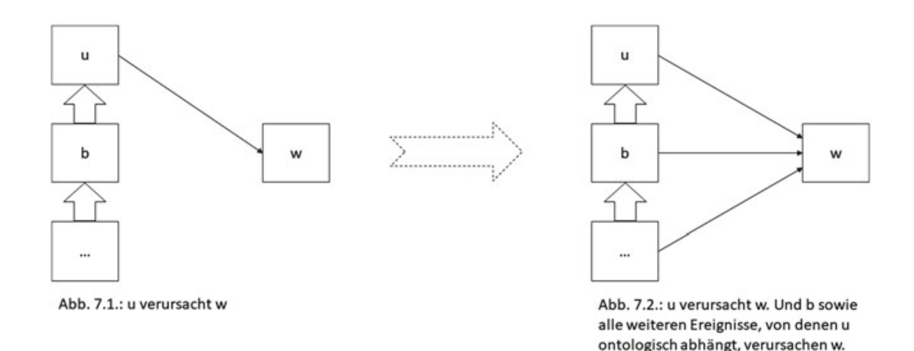


Abbildung 3.7 Prinzip der basalen Verursachung

In Abbildung 3.7 ist illustriert, welche Arten von Schlüssen das Prinzip der basalen Verursachung erlaubt. Die Situation in Abbildung 7.1. impliziert nach dem Prinzip der basalen Verursachung die Situation in Abbildung 7.2.. Abbildung 7.1. kann daher niemals eine vollständige kausale Struktur abbilden. Stattdessen ist die vollständige Struktur nur durch Abbildung 7.2. abgebildet.

Auch das Prinzip der basalen Verursachung hat Konsequenzen für die Möglichkeit reiner Intralevel-Verursachung. Wie das Prinzip der Abwärtsverursachung impliziert auch das Prinzip der basalen Verursachung, dass es eine reine Intralevel-Verursachung nur auf fundamentaler Ebene geben kann. Höherstufige Intralevel-Verursachung geht nach dem Prinzip der basalen Verursachung hingegen immer mit Aufwärtsverursachung einher: Wenn Susis Gedanke ihren Wunsch verursacht (mentale Intralevel-Verursachung), dann muss auch die neurologische Basis ihres Gedankens ihren Wunsch verursachen (neurologisch-mentale Aufwärtsverursachung). Eine weitere erwähnenswerte Konsequenz des

Prinzips der basalen Verursachung betrifft höherstufig-fundamentale Abwärtsverursachung. Diese impliziert nämlich nach dem Prinzip der basalen Verursachung fundamentale Intralevel-Verursachung: Wenn Susis Wunsch mikrophysikalische Veränderungen in den Elementarteilchen verursacht, die ihren motorischen Cortex konstituieren, dann werden diese mikrophysikalischen Veränderungen auch von der fundamentalen, $\text{physischen}_{\text{eng}}$ Basis ihres Wunsches verursacht.

Das Prinzip der basalen Verursachung kann auf eine ähnliche Weise vorläufig motiviert werden wie das Prinzip der Aufwärtsverursachung: Wenn u ontologisch abhängig von b ist, dann garantiert u, dass b eintritt. Wenn b aber hinreichende Ursache für w ist, dann garantiert u, dass w eintritt. Also garantiert b, dadurch, dass es das Eintreten von u garantiert, auch das Eintreten von w. Dies spricht dafür, dass b ebenfalls hinreichende Ursache für w ist.[64]

Ich habe die drei Prinzipien in diesem Abschnitt neutral hinsichtlich der Frage formuliert, ob sie die Relation der *hinreichenden* Verursachung oder die Relation der *Verursachung* betreffen. Diese Frage ist jedoch, wie noch deutlich werden wird, relevant für die Stärke der zu diskutierenden Erweiterungen des einfachen Exklusionsarguments. Ich werde diese Frage an späterer Stelle – in Abschnitt 4.2 – also wieder aufgreifen. Jedoch werde ich auch bei meiner Rekonstruktion der Erweiterungen des einfachen Exklusionsarguments vorerst weitgehend bei dieser Neutralität bleiben.

---

[64] Die erwähnten Probleme mit der vorläufigen Motivation des Prinzips der Aufwärtsverursachung treffen auch hier zu: Die Motivation wird bestenfalls für *hinreichende* Ursachen und *totale* Basen plausibel und beruht zunächst auf einer nomologischen oder regularitätstheoretischen Konzeption von Kausalität. Kim (1998: 43) gibt eine ähnliche Motivation (für eine Instanz) des Prinzips der basalen Verursachung: „If you take causation as grounded in nomological sufficiency, [b] qualifies as a cause of [w], for, since [b] is sufficient for [u] and [u] is sufficient for [w], [b] is sufficient for [w]". Ich habe die Buchstaben an die hiesige Verwendung angepasst. Vgl. auch Kim (2011: 127–128) für weitere knappe Begründungen im Rahmen anderer Theorien der Kausalität, auf die ich in Kapitel 4 noch detaillierter eingehe.

Schließlich lässt sich noch hinzufügen, dass das Prinzip der basalen Verursachung recht unmittelbar aus Shoemakers (2007) Auffassung von Realisierung zu folgen scheint. Denn „property P has Q as a realizer just in case (1) the forward-looking causal features of P are a subset of the forward-looking causal features of property Q and (2) the backward-looking causal features of P have as a subset the backward-looking causal features of Q" (Shoemaker (2007: 12)). Hier entspricht (1) dem Prinzip der basalen Verursachung und (2) dem Prinzip der Aufwärtsverursachung.

Auch Kroedel und Schulz (2016: 1915) leiten eine Entsprechung des Prinzips der basalen Verursachung aus ihrer Version eines Grounding-basierten Physikalismus ab: „Double Causation: Let m be a mental event, and let e be a physical effect of m. Then there is a physical event p that grounds m and causes e".

Die drei[65] Prinzipien der Verursachung im Stufenmodell der Realität eröffnen drei unterschiedliche Möglichkeiten, das einfache Exklusionsargument auszuweiten, so dass es nicht länger auf die fundamentalen, $\text{physischen}_{\text{eng}}$ (vermeintlichen) Wirkungen mentaler Ereignisse beschränkt bleibt. Das einfache Exklusionsargument dient dabei als Anker, von dem aus sich allgemeine Argumente gegen die Existenz mentaler Verursachung entwickeln lassen, wenn man eines der drei Prinzipien hinzuzieht. Im folgenden Abschnitt führe ich zunächst aus, wie das Prinzip der Abwärtsverursachung für eine Erweiterung des einfachen Exklusionsarguments genutzt werden kann.

3.3.6 Das Argument aus dem Prinzip der Abwärtsverursachung

Kommen wir also zurück zu der Frage, wie das einfache Exklusionsargument so ausgeweitet werden kann, dass es nicht länger ausschließlich fundamentale, $\text{physische}_{\text{eng}}$ Ereignisse betrifft. Eine erste Möglichkeit, die auch von Kim in Anspruch genommen wird[66], beruft sich auf das Prinzip der Abwärtsverursachung, das hier zur Erinnerung noch einmal erwähnt werden soll:

---

[65] Aufmerksamen Lesern wird die logische Möglichkeit eines vierten Prinzips nicht entgangen sein, das auf ähnliche Weise als Gegenstück zum Prinzip der basalen Verursachung dienen könnte, wie das Prinzip der Aufwärtsverursachung als Gegenstück zum Prinzip der Abwärtsverursachung dient. Ein solches Prinzip könnte wie folgt formuliert werden:

Viertes Prinzip: Wenn ein Ereignis u eine (hinreichende) Ursache für ein Ereignis w ist und ein Ereignis h ontologisch abhängig von u ist, dann ist auch h eine (hinreichende) Ursache für w.

Dieses Prinzip übernimmt im Kontext von Erweiterungen des einfachen Exklusionsarguments keine Funktion, weshalb ich an diese Stelle nicht weiter darauf eingehe. Kim hat eine Zeit lang ein verwandtes Prinzip im Rahmen einer möglichen Erwiderung auf Exklusionsargumente angewandt (vgl. Kim (1984b: 262)). In Kim (1989: 45–46) schildert er diese Idee wie folgt: „[A] mental event is a cause, or an effect, of another event in virtue of the fact that it is supervenient on some physical event standing in an appropriate causal relation to this event. Thus, mental properties are seen as deriving their causal potential from the physical properties on which they supervene.“ Vgl. Menzies (1988) für Kritik.

Ein ebenfalls verwandtes, aber direkt widersprechendes Prinzip wird in Abschnitt 3.3.9. noch einmal kurz Thema. Manche Exklusionsargumente beruhen nämlich auf einem Exklusionsprinzip, demzufolge gilt: Wenn ein Ereignis u eine (hinreichende) Ursache für ein Ereignis w ist und ein Ereignis h ontologisch abhängig von u ist, dann ist h *gerade keine* (hinreichende) Ursache für w. Vgl. z. B. List und Menzies (2009).

[66] Diese Erweiterung folgt im Prinzip Kims ‚Supervenience Argument', vgl. Kim (1993b: 203–211), Kim (1998: 38–47) und Kim (2005: 33–45).

Prinzip der Abwärtsverursachung: Wenn ein Ereignis u eine (hinreichende) Ursache für ein Ereignis w ist und Ereignis w ontologisch abhängig von einem Ereignis b ist, dann ist u auch eine (hinreichende) Ursache für Ereignis b.

Als zweite Zutat benötigen wir eine universelle Variante der These der ontologischen Abhängigkeit, die sich nicht auf mentale Ereignisse beschränkt, sondern alle höherstufigen Ereignisse betrifft:

Universale Ontologische Abhängigkeit: Jedes höherstufige Ereignis h ist ontologisch abhängig von einem $\text{physischen}_{\text{eng}}$ Ereignis p.

Mithilfe dieser beiden Thesen ist es nun einfach, in einem ersten Schritt zu zeigen, dass ein mentales Ereignis, insofern es ein höherstufiges Ereignis verursacht, auch ein fundamentales, $\text{physisches}_{\text{eng}}$ Ereignis verursacht. In einem zweiten Schritt können wir dann auf das einfache Exklusionsargument verweisen, um zu etablieren, dass das mentale Ereignis *nicht* ein fundamentales, $\text{physisches}_{\text{eng}}$ Ereignis verursacht. Hieraus können wir dann schließen, dass das mentale Ereignis auch keine höherstufigen Ereignisse verursacht.

Gehen wir also davon aus, dass m ein mentales Ereignis ist, das als Ursache für ein höherstufiges Ereignis h in Frage kommt. Aufgrund der These der ontologischen Abhängigkeit können wir postulieren, dass h eine $\text{physische}_{\text{eng}}$ Basis p* hat, von der h ontologisch abhängt. Aufgrund des Prinzips der Abwärtsverursachung muss m auch p* verursachen, wenn m h verursacht. Das ist also der erste Schritt dieser Erweiterung des einfachen Exklusionsarguments: *Wenn m h verursacht, dann verursacht m auch hs* $\textit{physische}_{\textit{eng}}$ *Basis p*.*

An dieser Stelle aber kommt das einfache Exklusionsargument ins Spiel: m kann p*, die $\text{physische}_{\text{eng}}$ Basis von h, nicht verursachen, weil p* aufgrund der These der kausalen Geschlossenheit bereits eine hinreichende, $\text{physische}_{\text{eng}}$ Ursache p hat. Aber m ist nach der These der Nicht-Identität nicht mit p (oder mit einem Teil von p) identisch. Aufgrund des einfachen Exklusionsprinzips wird m also als Ursache von p* ausgeschlossen. Das ist der zweite Schritt dieser Erweiterung des einfachen Exklusionsarguments: *m verursacht nicht hs* $\textit{physische}_{\textit{eng}}$ *Basis p*.*

Aus dem ersten Schritt und dem zweiten Schritt zusammengenommen folgt nun aber (nach modus tollens), dass m auch h nicht verursacht. Also: *m verursacht nicht h.*

Zur Verdeutlichung wieder eine Rekonstruktion des Arguments, das wir zur späteren Bezugnahme ‚das Argument aus dem Prinzip der Abwärtsverursachung' taufen können:

Erweiterung I: Das Argument aus dem Prinzip der Abwärtsverursachung
Es sei m ein mentales Ereignis. Es sei h ein höherstufiges Ereignis. p und p* seien zwei $\text{physische}_{\text{eng}}$ Ereignisse. Unter diesen Annahmen lässt sich das Argument aus dem Prinzip der Abwärtsverursachung wie folgt rekonstruieren:

Schritt 1:

(1) h ist ontologisch abhängig von p* (motiviert durch die These der ontologischen Abhängigkeit).
(2) Wenn m Ursache von h ist und h ontologisch abhängig von p* ist, dann ist m Ursache von p* (motiviert durch das Prinzip der Abwärtsverursachung).
(3) Also: Wenn m Ursache von h ist, dann ist m Ursache von p* (aus 1,2)

Schritt 2:

(4) p ist hinreichende Ursache für p* (motiviert durch die These der kausalen Geschlossenheit).
(5) m ist nicht mit p (oder einem Teil von p) identisch (motiviert durch die These der Nicht-Identität).
(6) Wenn p hinreichende Ursache von p* ist und m nicht mit p (oder einem Teil von p) identisch ist, dann ist m nicht Ursache von p* (motiviert durch das einfache Exklusionsprinzip).
(7) Also: m ist nicht Ursache von p* (aus 4, 5, 6).
(8) Also: m ist nicht Ursache von h (aus 3, 7).

‚h' kann in diesem Argument für ein beliebiges höherstufiges Ereignis stehen. Wir können sowohl neurologische Ereignisse als auch behaviorale Ereignisse oder mentale Ereignisse einsetzen. Durch diese Erweiterung ist die Limitierung des einfachen Exklusionsargument also überwunden: Das neue Argument betrifft nun nicht nur fundamentale, $\text{physische}_{\text{eng}}$ Ereignisse, sondern *alle* Ereignisse – ob fundamental oder höherstufig. Entsprechend lässt sich mit diesem Argument etablieren, dass Susis Gedanke nicht ihren Wunsch verursachen kann und ihr Wunsch nicht ihr Armheben, nicht die neurologischen Vorgänge in ihrem motorischen Cortex und nicht die mikrophysikalischen Veränderungen, die diese neurologischen Vorgänge konstituieren, verursachen kann. Mentale Ereignisse haben, wenn dieses Argument funktioniert, unter den Annahmen des nicht-reduktiven Physikalismus also *keinerlei Wirkungen.* Sie sind vollständig epiphänomenal.

3.3.7 Das Argument aus dem Prinzip der Aufwärtsverursachung

Wie ich in Abschnitt 2.3.2. ausgeführt habe, ergibt sich die eingeschränkte Reichweite des einfachen Exklusionsarguments aus der eingeschränkten Reichweite der These der kausalen Geschlossenheit. Ein weiterer Ansatz, diese Einschränkung aufzulösen besteht also darin, die Reichweite der These der kausalen Geschlossenheit auszuweiten. Hierfür kann man sich auf das Prinzip der Aufwärtsverursachung berufen, das hier zu Erinnerung noch einmal genannt wird:

> **Prinzip der Aufwärtsverursachung**: Wenn ein Ereignis u eine (hinreichende) Ursache für ein Ereignis w ist und ein Ereignis h ontologisch abhängig von Ereignis w ist, dann ist u auch eine (hinreichende) Ursache für Ereignis h.

Wieder ist es für dieses Argument nötig, eine universale Variante der These der ontologischen Abhängigkeit hinzuzuziehen:

> **Universale Ontologische Abhängigkeit**: Jedes höherstufige Ereignis h ist ontologisch abhängig von einem $\text{physischen}_{\text{eng}}$ Ereignis p.

Auf Grundlage dieser beiden Thesen lässt sich nun in einem ersten Schritt ein Argument konstruieren, das von der These der kausalen Geschlossenheit, die besagt, dass jedes *fundamentale,* $\textit{physische}_{eng}$ Ereignis eine hinreichende, $\text{physische}_{\text{eng}}$ Ursache hat, zu der Annahme führt, dass auch jedes *höherstufige* Ereignis eine hinreichende, $\text{physische}_{\text{eng}}$ Ursache hat. Und zwar wie folgt:

Es sei h ein beliebiges höherstufiges Ereignis. Auf Grundlage der These der ontologischen Abhängigkeit können wir postulieren, dass h eine $\text{physische}_{\text{eng}}$ Basis hat, die wir ‚p*‘ nennen können. Aufgrund der These der kausalen Geschlossenheit können wir zudem postulieren, dass p* eine hinreichende, $\text{physische}_{\text{eng}}$ Ursache p hat.[67] Das Prinzip der Aufwärtsverursachung erlaubt nun, zu schließen, dass p als hinreichende, $\text{physische}_{\text{eng}}$ Ursache von p* auch eine hinreichende, $\text{physische}_{\text{eng}}$ Ursache von h ist. Also hat auch h eine hinreichende, $\text{physische}_{\text{eng}}$ Ursache. Dasselbe gilt für jedes andere höherstufige Ereignis. Also hat jedes höherstufige Ereignis eine hinreichende, $\text{physische}_{\text{eng}}$ Ursache.[68]

---

[67] Auch hier versteckt sich wieder die Annahme, dass p* überhaupt eine hinreichende Ursache hat.

[68] Vgl. Esfeld (2005: 51), Baker (2009: 7) und Zhong (2011: 137) für Argumente, die nach diesem Schema funktionieren. Kim beruft sich in seiner Entwicklung von Exklusionsargumenten meines Wissen nirgendwo auf das Prinzip der Aufwärtsverursachung.

Als zweiten Schritt können wir uns nun wieder auf die Argumentation berufen, die aus dem einfachen Exklusionsargument bekannt ist. Allerdings können wir diese Argumentation nun direkt auf das höherstufige Ereignis h beziehen. Es sei m ein mentales Ereignis, das als Ursache für h in Frage kommt. h hat, wie im ersten Schritt gezeigt, aber bereits eine hinreichende $\text{physische}_{\text{eng}}$ Ursache p. m ist nach der These der Nicht-Identität nicht mit p (oder einem Teil von p) identisch. Das einfache Exklusionsprinzip schließt daher aus, dass m ebenfalls h verursacht. Also verursacht m nicht h.

Zur späteren Bezugnahme nenne ich dieses Argument ‚das Argument aus dem Prinzip der Aufwärtsverursachung' und führe zur Verdeutlichung die folgende Rekonstruktion ein:

Erweiterung II: Das Argument aus dem Prinzip der Aufwärtsverursachung
Es sei m ein mentales Ereignis. Es sei h ein höherstufiges Ereignis. p und p* seien zwei $\text{physische}_{\text{eng}}$ Ereignisse. Unter diesen Annahmen lässt sich das Argument aus dem Prinzip der Aufwärtsverursachung wie folgt rekonstruieren:

Schritt 1:

(1) h ist ontologisch abhängig von p* (motiviert durch die These der ontologischen Abhängigkeit).
(2) p ist hinreichende Ursache von p* (motiviert durch die These der kausalen Geschlossenheit).
(3) Wenn p hinreichende Ursache für p* ist und h ontologisch abhängig von p* ist, dann ist p hinreichende Ursache für h.
(4) Also: p ist hinreichende Ursache für h (aus 1, 2, 3).

Schritt 2:

(5) m ist nicht mit p (oder einem Teil von p) identisch (motiviert durch die These der Nicht-Identität).
(6) Wenn p hinreichende Ursache von h ist und m nicht mit p (oder einem Teil von p) identisch ist, dann ist m nicht Ursache von h (motiviert durch das einfache Exklusionsprinzip)
(7) Also: m ist nicht Ursache von h (aus 4, 5, 6).

Auch das Argument aus dem Prinzip der Aufwärtsverursachung ist im Gegensatz zum einfachen Exklusionsargument nicht auf $\text{physische}_{\text{eng}}$ Ereignisse eingeschränkt. Ebenso wie das Argument aus der Abwärtsverursachung erlaubt also auch dieses Argument, eine uneingeschränkt allgemeine epiphänomenalistische

Konsequenz zu begründen: Wenn das Argument funktioniert, haben mentale Ereignisse im nicht-reduktiven Physikalismus keinerlei Wirkungen und sind vollständig epiphänomenal.[69]

3.3.8 Das Argument aus dem Prinzip der basalen Verursachung

Auch aus dem Prinzip der basalen Verursachung ergibt sich ein Exklusionsargument, das die Beschränkung der Reichweite des einfachen Exklusionsarguments vermeidet. Zur Erinnerung sei das Prinzip der basalen Verursachung hier noch einmal genannt:

> **Prinzip der basalen Verursachung**: Wenn ein Ereignis u eine (hinreichende) Ursache für ein Ereignis w ist und Ereignis u ontologisch abhängig von einem Ereignis b ist, dann ist auch Ereignis b eine (hinreichende) Ursache für Ereignis w.

Wie auch im Argument aus dem Prinzip der Abwärtsverursachung und dem Argument aus dem Prinzip der Aufwärtsverursachung spielt in dem in diesem Abschnitt zu entwickelnden Argument aus dem Prinzip der basalen Verursachung die These der ontologischen Abhängigkeit eine wichtige Rolle. Hierbei ist es jedoch nicht notwendig, eine universale These der ontologischen Abhängigkeit

---

[69] Der folgende Unterschied zwischen dem Argument aus dem Prinzip der Aufwärtskausalität und dem Argument aus dem Prinzip der Abwärtskausalität sei an dieser Stelle noch genannt: Während das Argument aus dem Prinzip der Aufwärtskausalität auf dem Weg zum epiphänomenalistischen Argumentationsziel etabliert, dass alle Ereignisse hinreichende $\text{physische}_{\text{eng}}$ Ursachen haben, etabliert das Argument aus dem Prinzip der Abwärtskausalität keine solche These. Das Argument aus dem Prinzip der Abwärtskausalität ist also im Gegensatz zum Argument aus dem Prinzip der Aufwärtskausalität damit kompatibel, dass ausschließlich $\text{physische}_{\text{eng}}$ Ereignisse hinreichende $\text{physische}_{\text{eng}}$ Ursachen haben. Höherstufige Ereignisse können in Übereinstimmung mit dem Argument aus dem Prinzip der Abwärtskausalität auch gar keine Ursachen haben.

Man könnte daher sagen, dass das Argument aus dem Prinzip der Aufwärtskausalität einen klassischen Epiphänomenalismus impliziert, demzufolge mentale Ereignisse zwar verursacht werden, aber selbst keine Wirkungen haben. Das Argument aus dem Prinzip der Abwärtskausalität ist hingegen mit einer alternativen Version des Epiphänomenalismus kompatibel, der zufolge mentale Ereignisse ganz aus der kausalen Struktur der Welt herausfallen, da sie weder verursacht werden noch etwas verursachen. Vgl. auch Walter (2009) für diese Unterscheidung zwischen zwei Versionen des Epiphänomenalismus.

vorauszusetzen. Stattdessen reicht für die Zwecke des Arguments aus dem Prinzip der basalen Verursachung eine auf mentale Ereignisse eingeschränkte Version der These der ontologischen Abhängigkeit:

> **Mentale Ontologische Abhängigkeit**: Jedes mentale Ereignis m ist ontologisch abhängig von einem physischen_{eng} Ereignis p.

Das Prinzip der basalen Verursachung und die These der ontologischen Abhängigkeit haben in Kombination die folgende Konsequenz: Wenn ein mentales Ereignis m eine beliebige (fundamentale oder höherstufige) Wirkung w hat, dann hat w auch eine (hinreichende) physische_{eng} Ursache – nämlich die physische_{eng} Basis von m. Dies ist der erste Schritt des Arguments: *Wenn m eine (hinreichende) Ursache für w ist, dann ist auch p, die* $\textit{physische}_{eng}$ *Basis von m, eine (hinreichende) Ursache für w.*

Der zweite Schritt des Arguments besteht nun wieder in einer Überlegung, die bereits aus dem einfachen Exklusionsargument bekannt ist: Gegeben, dass m nach der These der Nicht-Identität nicht mit p (oder einem Teil von p) identisch ist, und dass w nach dem einfachen Exklusionsprinzip nicht mehr als eine (hinreichende) Ursache hat, kann w nicht von m *und* p verursacht werden. Dies ist der zweite Schritt des Arguments: *Wenn p eine (hinreichende) Ursache für w ist, dann ist m keine (hinreichende) Ursache für w.*

Die beiden Schritte haben zusammengenommen die paradoxe Konsequenz, dass aus der Annahme, dass m (hinreichende) Ursache für w ist, folgt, dass m *nicht* (hinreichende) Ursache für w ist. Die Konklusion dieses Arguments ist also nicht bloß, dass mentale Ereignisse im nicht-reduktiven Physikalismus keine Wirkungen haben. Vielmehr wird gezeigt, dass die Annahme, dass mentale Ereignisse Wirkungen haben, unter Voraussetzung des nicht-reduktiven Physikalismus *selbstwiderlegend* ist: Sie führt in einen Widerspruch.[70]

Wieder biete ich zur Klärung eine Rekonstruktion des Arguments an:

---

[70] Menzies (2013: 61–65) entwickelt ein ähnliches Argument. Das Argument könnte alternativ auch als *reductio ad absurdum* konstruiert werden: Aus der Annahme, dass m w verursacht, können wir einen Widerspruch ableiten. Vgl. Menzies (2013: 65). Auch Zhong (2011: 131–135) entwickelt ein verwandtes Argument.

Kims Verwendung des Prinzips der basalen Verursachung im Rahmen des ‚Supervenience Argument ‘(vgl. Kim (1998: 43f) und Kim (2005: 41)) spielt hingegen eine andere Rolle. Denn Kim bezieht sich in diesem Argument nicht nur auf das Prinzip der basalen Verursachung, sondern zusätzlich auf die These der kausalen Geschlossenheit und auf das Prinzip der Abwärtsverursachung. Tatsächlich scheint mir Kims Verwendung des Prinzips der basalen Verursachung im ‚Supervenience Argument‘ redundant zu sein.

Erweiterung III: Das Argument aus dem Prinzip der basalen Verursachung
Es sei m ein mentales Ereignis. Es sei w ein beliebiges Ereignis. p und p* seien zwei physische$_{eng}$ Ereignisse. Unter diesen Annahmen lässt sich das Argument aus dem Prinzip der basalen Verursachung wie folgt rekonstruieren:

Schritt 1:

(1) m ist ontologisch abhängig von p (motiviert durch die These der ontologischen Abhängigkeit).
(2) Wenn m (hinreichende) Ursache von w ist und m ontologisch abhängig von p ist, dann ist p (hinreichende) Ursache von w (motiviert durch die These der basalen Verursachung).
(3) Wenn m (hinreichende) Ursache von w ist, dann ist p (hinreichende) Ursache von w (aus 1, 2).

Schritt 2:

(4) m ist nicht mit p (oder einem Teil von p) identisch (motiviert durch die These der Nicht-Identität).
(5) Wenn p hinreichende Ursache von w ist und m nicht mit p (oder einem Teil von p) identisch ist, dann ist m nicht (hinreichende) Ursache von w (motiviert durch das einfache Exklusionsprinzip).
(6) Also: Wenn p (hinreichende) Ursache von w ist, dann ist m nicht (hinreichende) Ursache von w (aus 4, 5).
(7) Also: Wenn m (hinreichende) Ursache von w ist, dann ist m nicht (hinreichende) Ursache von w (aus 3, 6).

Das Argument aus dem Prinzip der basalen Verursachung ist ebenso wie das Argument aus dem Prinzip der Abwärtsverursachung und das Argument aus dem Prinzip der Aufwärtsverursachung uneingeschränkt allgemein. Es betrifft jede vermeintliche Wirkung eines mentalen Ereignisses. Eine Einschränkung auf fundamentale physische$_{eng}$ Ereignisse liegt nicht vor.

Eine Besonderheit dieses Arguments ist, dass es sich an keiner Stelle auf die These der kausalen Geschlossenheit beruft. Auch sind keine empirischen Evidenzen dafür nötig, dass w *tatsächlich* eine hinreichende, physische$_{eng}$ Ursache hat. Das Argument würde auch dann noch funktionieren, wenn kein einziges Ereignis eine hinreichende, physische$_{eng}$ Ursache hätte. Der Grund dafür ist, dass die Konklusion des Arguments wie auch seine Prämissen konditional sind: *Wenn* m eine (hinreichende) Ursache für w ist, dann ist auch p eine (hinreichende) Ursache für w. Und: *Wenn* p eine (hinreichende) Ursache für w ist, dann ist m keine (hinreichende) Ursache für w. Also: *Wenn* m eine hinreichende Ursache für w ist, dann ist m keine hinreichende Ursache für w. Für die Wahrheit dieser konditionalen Aussagen ist es

nicht nötig, dass m oder p *tatsächlich* w verursachen. Daher ist es auch nicht nötig, die These der kausalen Geschlossenheit einzubeziehen, um zu etablieren, dass p *tatsächlich* w verursacht. Es reicht vielmehr die Argumentation dafür, dass w eine hinreichende, $physische_{eng}$ Ursache hat, *insofern w von m (hinreichend) verursacht wird*. Denn schon hieraus lässt sich unter Hinzunahme der These der Nicht-Identität und des einfachen Exklusionsprinzips die Konklusion ableiten, dass die Annahme der Existenz mentaler Verursachung selbstwiderlegend ist.

Kim argumentiert jedoch in einem neueren Aufsatz dafür, dass die These der kausalen Geschlossenheit unter Voraussetzung der These der ontologischen Abhängigkeit aus dem Prinzip der basalen Verursachung folgt.[71] Das Argument ist – übertragen auf die hier verwendete Terminologie und in manchen Hinsichten ergänzt – in groben Zügen dieses:

Ereignisse sind entweder höherstufig oder fundamental und im engen Sinne physisch. Höherstufige Ereignisse haben aufgrund der These der ontologischen Abhängigkeit fundamentale, $physische_{eng}$ Basen. Wenn ein Ereignis nun eine hinreichende höherstufige Ursache hat, dann hat es aufgrund des Prinzips der basalen Verursachung auch eine hinreichende, $physische_{eng}$, fundamentale Ursache – nämlich die physische Basis der höherstufigen Ursache. Wenn ein Ereignis hingegen keine hinreichende, höherstufige Ursache hat, dann hat es (aufgrund der erschöpfenden Unterscheidung zwischen höherstufigen und fundamentalen Ereignissen) entweder überhaupt keine hinreichende Ursache oder eine hinreichende fundamentale, $physische_{eng}$ Ursache. Wenn ein Ereignis also überhaupt eine hinreichende Ursache hat, dann hat es eine hinreichende fundamentale $physische_{eng}$ Ursache. Die These der kausalen Geschlossenheit ist etabliert.[72]

Was durch diese Argumentation jedoch nicht etabliert wird, ist, dass die These der kausalen Geschlossenheit gehaltvoll ist. Auch dieses Argument würde sogar dann funktionieren, wenn überhaupt kein Ereignis eine hinreichende $physische_{eng}$ Ursache hat, so dass die These der kausalen Geschlossenheit leer wahr ist. Die anderen beiden Erweiterungen des einfachen Exklusionsarguments setzen natürlich eine gehaltvolle These der kausalen Geschlossenheit voraus. Das Argument aus dem Prinzip der basalen Verursachung tut dies nicht (und impliziert auch keine solche These).

Für den vorliegenden Kontext ist aber ohnehin nur dies relevant: Das Argument aus dem Prinzip der basalen Verursachung muss sich im Gegensatz zu den anderen diskutierten Exklusionsargumenten nicht explizit auf die These der kausalen Geschlossenheit beziehen und muss diese These insbesondere nicht in Anspruch

[71] Vgl. Kim (2011: 127–128).

[72] Vgl. auch Morris (2014) für weitere Diskussion.

nehmen, um zu etablieren, dass irgendein Ereignis tatsächlich eine hinreichende physische$_{eng}$ Ursache hat.

Dass das Argument aus dem Prinzip der basalen Verursachung auf die These der kausalen Geschlossenheit verzichtet, wirft eine interessante Frage auf. Denn im einfachen Exklusionsargument spielt die These der kausalen Geschlossenheit eine tragende Rolle: Sie sorgt für die Asymmetrie zwischen den zwei Kandidaten für Ursachen von p*. Da das Exklusionsprinzip lediglich vorschreibt, dass nicht sowohl p als auch m (hinreichende) Ursachen von p* sein können, und daher kompatibel damit ist, dass m p* verursacht und p als Ursache ausgeschlossen wird, muss die These der kausalen Geschlossenheit hinzugezogen werden, um zu etablieren, dass m als Ursache von p* ausgeschlossen wird.[73] Es stellt sich nun die Frage, wie eine solche Asymmetrie eigentlich in Bezug auf das Argument aus dem Prinzip der basalen Verursachung zustande kommt. Was sorgt dafür, dass die Annahme, dass m eine (hinreichende) Ursache für w ist, selbstwiderlegend ist, aber die Annahme, dass p eine (hinreichende) Ursache für w ist, nicht selbstwiderlegend ist?

Das Exklusionsprinzip spielt für diese Asymmetrie tatsächlich keine Rolle. Es ist, wie Kim schreibt, „neutral with respect to the mental-physical competition“[74]. Die Asymmetrie ergibt sich daher nicht aus dem zweiten Schritt des Arguments aus dem Prinzip der basalen Verursachung. Das sieht man schon daran, dass die Kontraposition der Konklusion des zweiten Schritts („(6) Wenn p (hinreichende) Ursache von w ist, dann ist m nicht (hinreichende) Ursache von w“) lautet: Wenn m (hinreichende) Ursache für w ist, dann ist p nicht (hinreichende) Ursache für w. Die Asymmetrie muss also im ersten Schritt des Arguments zu finden sein.

Und hier ist sie zu finden: Die Asymmetrie zwischen m und p ergibt sich gewissermaßen aus der Asymmetrie des Prinzips der basalen Verursachung sowie der Asymmetrie der Relation der ontologischen Abhängigkeit. Das Prinzip der basalen Verursachung erlaubt nämlich, aus ‚m verursacht w‘ und ‚m ist ontologisch abhängig von p‘ zu schließen: ‚p verursacht w‘. Aber es erlaubt nicht, aus ‚p verursacht w‘ und ‚m ist ontologisch abhängig von p‘ zu schließen: ‚m verursacht w‘. Deswegen können wir im ersten Schritt des Arguments auch nicht ableiten: *Wenn p w verursacht, dann verursacht auch m w*. Zwar stünde diese These in Konflikt mit der These der Nicht-Identität und dem Exklusionsprinzip. Jedoch kann sie nicht auf Grundlage des Prinzips der basalen Verursachung begründet werden. Deshalb ist nur die Annahme, dass m eine (hinreichende) Ursache von w ist, selbstwiderlegend, während die Annahme, dass p eine (hinreichende) Ursache für w ist, konsistent bleibt.

[73] Vgl. wiederum Kim (2005: 43).

[74] Kim (2005: 43).

Das Argument aus dem Prinzip der basalen Verursachung nimmt eine Sonderrolle innerhalb der drei bisher diskutierten Erweiterungen des einfachen Exklusionsarguments ein. Denn es weicht aufgrund des Verzichts auf die These der kausalen Geschlossenheit sowie mit Blick auf die Konklusion, dass mentale Verursachung selbstwiderlegend ist, stärker von dem einfachen Exklusionsargument ab, als das Argument aus dem Prinzip der Abwärtsverursachung und das Argument aus dem Prinzip der Aufwärtsverursachung.

Eine Gemeinsamkeit mit den anderen Erweiterungen besteht hingegen darin, dass es nicht nur $\text{physische}_{\text{eng}}$ (vermeintliche) Wirkungen mentaler Ereignisse betrifft, sondern *alle* (vermeintlichen) Wirkungen mentaler Ereignisse. Die drei Erweiterungen sind also verschiedene Möglichkeiten, die Limitierung des einfachen Exklusionsarguments zu vermeiden.

Eine zweite Gemeinsamkeit besteht darin, dass alle drei Argumente sich auf die These der ontologischen Abhängigkeit berufen. Sie sind also – im Gegensatz zum einfachen Exklusionsargument, das auch dualistische Positionen angreift – spezifisch auf den nicht-reduktiven Physikalismus gerichtet: Sie nehmen die Kernthesen des nicht-reduktiven Physikalismus als Ausgangspunkt, um (unter Hinzunahme weiter Prinzipien) aufzuzeigen, dass mentale Ereignisse keine Wirkungen haben.[75] Auf diese Weise leisten sie den für den ersten Schritt des Vorwurfs des Epiphänomenalismus zentralen Übergang vom nicht-reduktiven Physikalismus zum Epiphänomenalismus: *Wenn der nicht-reduktive Physikalismus wahr ist, dann gibt es keine mentale Verursachung.*

3.3.9 Das einfache Exklusionsprinzip und genuine Überdetermination

Eine weitere Gemeinsamkeit aller bis hierher diskutierten Exklusionsargumente besteht darin, dass sich alle Argumente auf das einfache Exklusionsprinzip berufen. Im Folgenden werde ich dieses Prinzip daher genauer in den Blick nehmen. In Abschnitt 3.3.1. habe ich das einfache Exklusionsprinzip wie folgt formuliert:

> **Das einfache Exklusionsprinzip**: Wenn ein Ereignis eine hinreichende Ursache zu t hat, dann hat es darüber hinaus keine weitere Ursache zu t.

[75] Kim (1998: 30) schreibt: „[T]he exclusion problem is distinctive in that it strikes at the very heart of physicalism.“ Diese Bemerkung trifft tatsächlich auf die Erweiterungen des einfachen Exklusionsarguments eher zu als auf das einfache Exklusionsargument selbst.

Schauen wir uns noch einmal die Funktion des Exklusionsprinzips in den bisher besprochenen Exklusionsargumenten an. Das Exklusionsprinzip kommt immer dann zur Anwendung, wenn wir zwei verschiedene Kandidaten für (hinreichende) Ursachen für ein gegebenes Ereignis haben: Im einfachen Exklusionsargument und im Argument aus dem Prinzip der Abwärtskausalität garantiert das Exklusionsprinzip, dass das $\text{physische}_{\text{eng}}$ Ereignis p* nicht sowohl von der hinreichenden Ursache p als auch von dem mentalen Ereignis m verursacht wird. Im Argument aus dem Prinzip der Aufwärtskausalität sichert das Exklusionsprinzip, dass das höherstufige Ereignis h nicht sowohl eine hinreichende $\text{physische}_{\text{eng}}$ Ursache als auch eine mentale Ursache hat. Im Argument aus dem Prinzip der basalen Verursachung stellt das Exklusionsprinzip sicher, dass ein beliebiges Ereignis w nicht sowohl von m als auch von p, der $\text{physischen}_{\text{eng}}$ Basis von m, verursacht wird. In allen Argumenten würde eine Ablehnung des Exklusionsprinzips dazu führen, dass die in Frage stehende Wirkung eine hinreichende physische Ursache und gleichzeitig eine zusätzliche mentale Ursache hat.

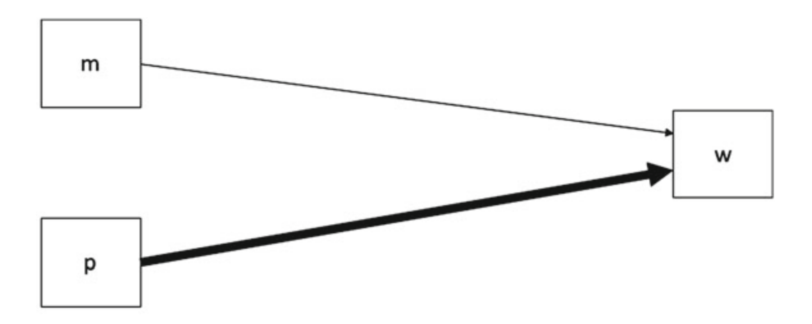
Eine Ablehnung des Exklusionsprinzips würde also in jedem dieser Fälle zu der Option führen, dass die Wirkungen mentaler Ereignisse in einem gewissen Sinne *überdeterminiert* sind. Sie würde zu einem *Modell von mentaler Verursachung als Überdetermination* führen (siehe Abbildung 3.8). Das einfache Exklusionsprinzip hat also in allen Exklusionsargumenten die Funktion, das Modell von mentaler Verursachung als Überdetermination als Rückzugsort nicht-reduktiver Physikalist*innen auszuschließen.

Was bedeutet es aber genauer, dass ein Ereignis überdeterminiert ist? Für die vorliegenden Zwecke können wir die folgende, allgemeine Definition von Überdetermination voraussetzen:

> **Überdetermination**: Ein Ereignis w ist genau dann überdeterminiert, wenn gilt: w hat eine hinreichende Ursache zu t und darüber hinaus eine weitere Ursache zu t.[76]

Es ist eindeutig, dass das Modell der mentalen Verursachung als Überdetermination impliziert, dass die Wirkungen mentaler Ereignisse in diesem Sinne

[76] Gelegentlich finden sich deutlich anspruchsvollere Definitionen von Überdetermination (vgl. z. B. Carey (2011: 256): „e is overdetermined $=_{\text{Df}}$ e has two simultaneous sufficient causes that are metaphysically, nomologically, and causally independent“ Carey spielt in diesem Aufsatz auch eine Reihe weiterer möglicher Definitionen durch). In den Begrifflichkeiten, die ich hier voraussetze und in den kommenden Absätzen ausführe, sind solche anspruchsvolleren Definitionen als Definitionen von *genuiner* Überdetermination zu verstehen. Vgl. auch Bennett (2003: 476).



Erläuterung: Der dicke Pfeil steht für die Relation der hinreichenden Verursachung. Der dünnere Pfeil steht für die Relation der (hinreichenden) Verursachung.
p ist ein physisches$_{eng}$ Ereignis. m ist ein mentales Ereignis. w ist eine beliebige Wirkung von m.

Abbildung 3.8 Mentale Verursachung als Überdetermination

überdeterminiert sind: w hat eine hinreichende physische Ursache p und darüber hinaus eine gleichzeitige mentale Ursache m.

Das einfache Exklusionsprinzip schließt das Modell der mentalen Verursachung als Überdetermination aus, indem es in uneingeschränkter Allgemeinheit ausschließt, dass ein Ereignis eine hinreichende Ursache und darüber hinaus eine zeitgleiche, zusätzliche Ursache hat. Damit schießt das einfache Exklusionsprinzip jedoch nach allgemeiner Auffassung über das Ziel hinaus. Denn es schließt nicht *nur* das Modell mentaler Verursachung als Überdetermination aus. Stattdessen schließt es jede Form von Überdetermination aus. Wenn das einfache Exklusionsprinzip also wahr ist, dann gibt es keine Überdetermination.

Dass dies eine sehr starke – vielleicht *zu starke* – Behauptung ist, sieht man an dem folgenden Beispiel:

Genuine Überdetermination – zwei Attentäter: Zwei Attentäter schießen zeitgleich und aus gleichem Abstand unabhängig voneinander auf ein Opfer, so dass ihre Kugeln zeitgleich eintreffen. Das Opfer stirbt. Jeder einzelne Schuss hätte aber auch in Abwesenheit des anderen Schusses zum Tod des Opfers geführt. In diesem Fall hat der Tod

des Opfers zwei hinreichende Ursachen: Jeder der beiden Schüsse ist hinreichende Ursache für den Tod des Opfers. Der Tod des Opfers ist überdeterminiert.[77]

Wenn das einfache Exklusionsprinzip wahr ist, sind solche Fälle von genuiner Überdetermination ebenso ausgeschlossen wie das Modell von mentaler Verursachung als Überdetermination. Genuine Überdetermination scheint aber nicht allgemein ausgeschlossen zu sein: Wirkungen sind zwar nicht im Regelfall überdeterminiert und Fälle von genuiner Überdetermination sind seltene Ausnahmen. Aber ein Prinzip, das die Existenz genuiner Überdetermination ganz allgemein ablehnt, scheint auf den ersten Blick nicht plausibel. Der Fall der zwei Attentäter ist also ein Gegenbeispiel gegen das einfache Exklusionsprinzip.

Nun ließe sich das einfache Exklusionsprinzip natürlich verteidigen, indem gezeigt wird, dass es dem ersten Anschein zum Trotz doch keine Fälle von genuiner Überdetermination gibt.

Es ließe sich etwa argumentieren, dass sich alle vermeintlichen Fälle genuiner Überdetermination bei genauerer Analyse als Fälle von gemeinschaftlicher Verursachung entpuppen: Keine der beiden Ursachen ist für sich genommen hinreichend für die Wirkung. Tatsächlich hätte der Schuss des ersten Attentäters in Abwesenheit des Schusses des zweiten Attentäters nämlich nicht *genau diesen* Tod des Opfers verursacht. Hätte nur ein Schuss stattgefunden, wäre das Opfer zwar auch gestorben. Jedoch wäre es auf eine andere Art und Weise gestorben und es hätte also ein anderes Ereignis stattgefunden. Nach diesem Ansatz garantieren nur beide Schüsse in Kombination, dass es zu *diesem* spezifischen Tod des Opfers kommt. Es liegen also gar nicht zwei hinreichende Ursachen für ein und dieselbe Wirkung vor, sondern zwei Teilursachen.[78]

Alternativ könnte man dafür argumentieren, dass die richtige Analyse des Falls der zwei Attentäter die beiden Schüsse nicht jeweils einzeln als zwei hinreichende Ursachen klassifiziert, sondern nur die ‚Summe‘[79] der beiden Schüsse als eine einzige hinreichende Ursache. Entgegen dem ersten Anschein hat der Tod dann

---

[77] Das Beispiel ist sehr verbreitet und gehört gewissermaßen zur Folklore der Debatte um Verursachung, Überdetermination und mentale Verursachung. Vgl. z. B. Loewer (2002: 657), Kim (2005: 48), Bernstein (2017: 165) und zahlreiche Weitere.

[78] Vgl. Bunzl (1979), dessen Argumentation auf der Annahme beruht, dass Ereignisse maximal modal fragil individuiert werden: Schon kleinste mögliche Änderungen an einem Ereignis ergeben ein zweites Ereignis. Vgl. Funkhouser (2002: 339) und Schaffer (2003b: 26f) für Kritik.

[79] Was hier mit ‚Summe‘ gemeint sein kann, müsste natürlich weiter spezifiziert werden. Schaffer (2003b) geht in seiner kritischen Diskussion von einer mereologischen Summe der beiden Ereignisse aus, die ein weiteres Ereignis bildet.

nicht zwei hinreichende Ursachen (den ersten Schuss und den zweiten Schuss), sondern nur eine einzige (die ‚Summe' aus dem ersten und zweiten Schuss). Allgemein sind (vermeintliche) Fälle von genuiner Überdetermination dann immer als Fälle von einfacher hinreichender Verursachung durch ein zusammengesetztes Ereignis zu interpretieren.[80]

Wie vielversprechend diese Strategien zur Vermeidung von Fällen genuiner Überdetermination sind, sei hier dahingestellt. Wichtig ist an dieser Stelle ohnehin lediglich Folgendes: Wenn Exklusionsargumente unter Berufung auf das einfache Exklusionsprinzip formuliert werden, können sie zurückgewiesen werden, indem die prinzipielle Möglichkeit von Fällen von genuiner Überdetermination aufgezeigt wird. Nicht-reduktive Physikalist*innen können also alle bis hierher diskutierten Exklusionsargumente abwehren, indem sie für die Möglichkeit des Falls der zwei Attentäter argumentieren.[81] Dieser Umstand ist aus Sicht der Kritiker*innen des nicht-reduktiven Physikalismus nicht wünschenswert. Und es ist auch nicht zu sehen, warum Kritiker*innen des nicht-reduktiven Physikalismus sich auf die Unmöglichkeit genuiner Überdetermination festzulegen hätten. Aus diesem Grund werden Exklusionsargumente für gewöhnlich nicht unter Voraussetzung des einfachen Exklusionsprinzips formuliert.

In weiten Teilen der Debatte hat sich stattdessen ein Exklusionsprinzip durchgesetzt, das eine explizite Ausnahmeklausel für Fälle von genuiner Überdetermination enthält:

> **Das übliche Exklusionsprinzip**: Wenn ein Ereignis eine hinreichende Ursache zu t hat, dann hat es darüber hinaus keine weitere Ursache zu t – es sei denn, es handelt sich um einen Fall von genuiner Überdetermination.[82]

---

[80] Vgl. Schaffer (2003b), der eine entsprechende Unterscheidung zwischen individualistischer und kollektivistischer Überdetermination trifft. Individualistische Überdetermination liegt vor, wenn zwei einzelne Ereignisse jeweils als hinreichende Ursachen für dieselbe Wirkung gezählt werden, während bei kollektivistischer Überdetermination nur die Summe der beiden Ereignisse als hinreichende Ursache gezählt wird. Ich unterstelle in meiner Diskussion, dass nur individualistische Überdetermination als genuine Überdetermination zählt. Hintergrund dieser Annahme ist, dass nur aus individualistischer Überdetermination ein Gegenbeispiel gegen das einfache Exklusionsprinzip zu gewinnen ist.

[81] Vgl. Auch Kim (1998: 53) und Morris (2019: 65).

[82] Vgl. Kim (2005: 17): „If an event e has a sufficient cause c at t, no event at t distinct from c can be a cause of e (unless this is a genuine case of causal overdetermination)". Vgl. auch Bennett (2003: 472), Bennett (2008: 281), Árnadóttir und Crane (2013: 253), Morris (2019: 64) und zahlreiche Weitere.

Dieses Prinzip kann nun nicht mehr durch einen Verweis auf Fälle genuiner Überdetermination zurückgewiesen werden. Allerdings kann es auch für sich genommen nicht die Funktion übernehmen, die das einfache Exklusionsprinzip in den bisherigen Exklusionsargumenten übernommen hat. Wir können beispielsweise im einfachen Exklusionsargument nicht aufgrund des üblichen Exklusionsprinzips schließen, dass m keine Ursache für p* ist. Denn p* könnte *genuin überdeterminiert* sein. In diesem Fall wäre das übliche Exklusionsprinzip nicht auf p* anwendbar. Um zur gewünschten Konklusion zu gelangen, müssen wir uns also auf eine zusätzliche Annahme stützen: p* ist *nicht* genuin überdeterminiert. Entsprechend müssen auch die diskutierten Erweiterungen des einfachen Exklusionsarguments modifiziert werden, wenn sie sich statt auf das einfache Exklusionsprinzip auf das übliche Exklusionsprinzip beziehen: Es muss jeweils angenommen werden, dass die in Frage stehende vermeintliche Wirkung des mentalen Ereignisses nicht *genuin* überdeterminiert ist – dass sie also nicht im selben Sinne überdeterminiert ist, wie der Tod des Opfers im Fall der zwei Attentäter.

Diese Überlegung legt nahe, dass Exklusionsargumente, die sich auf das übliche Exklusionsprinzip stützen, die folgende, allgemeine Annahme voraussetzen:

> **Überdeterminationsverbot**: Die Wirkungen mentaler Ereignisse sind nicht genuin überdeterminiert.[83]

Die Idee hinter dieser Annahme ist dann, dass sich mentale Verursachung als Überdetermination, wie sie im Rahmen des nicht-reduktiven Physikalismus nahegelegt wird, bedeutend von der *genuinen* Überdetermination unterscheidet, die ihren Ausdruck im Beispiel der zwei Attentäter findet. Aus diesem Grund findet das übliche Exklusionsprinzip Anwendung auf vermeintliche Fälle von mentaler Verursachung als Überdetermination. Und unter dieser Voraussetzung schließt das übliche Exklusionsprinzip das Modell der mentalen Verursachung als Überdetermination aus.[84]

---

[83] Vgl. z. B. Bennett (2008: 281): „The effects of mental causes are not systematically overdetermined; they are not on a par with the deaths of firing squad victims“.

[84] Das Überdeterminationsverbot ist vermutlich nicht in voller Allgemeinheit wahr. Denn wenn der Tod eines Opfers in seltenen Fällen genuin überdeterminiert sein kann, dann kann sicherlich auch eine vermeintliche Wirkung eines mentalen Ereignisses wie Susis Armbewegung in seltenen Fällen genuin überdeterminiert sein. Nehmen wir etwa an, Susis Sitznachbarin hebt Susis Arm just in dem Moment an, in dem ihr Arm auch aufgrund der physischen Basis ihres Wunsches in Bewegung gerät. Ihre Armbewegung hat dann zwei unabhängige hinreichende physische Ursachen und ist genuin überdeterminiert.

Eigentlich müsste man das Prinzip also wohl einschränken:

Das übliche Exklusionsprinzip setzt also eine begriffliche Unterscheidung zwischen *genuiner Überdetermination* auf der einen Seite und *nicht-genuiner Überdetermination* auf der anderen Seite voraus. Genuine Überdetermination ist selten, aber (womöglich)[85] nicht komplett ausgeschlossen. Nicht-genuine Überdetermination ist hingegen aufgrund des üblichen Exklusionsprinzips allgemein ausgeschlossen.

In beiden Fällen haben wir eine hinreichende Ursache und gleichzeitig eine zusätzliche Ursache. Sowohl genuine als auch nicht-genuine Überdetermination sind also Fälle von Überdetermination nach der obigen Definition. Dies wirft die Frage auf, worin sich genuine Überdetermination von nicht-genuiner Überdetermination unterscheidet. Hierzu schreibt Kim Folgendes:

> In standard cases of overdetermination, like two bullets hitting the victim's heart at the same time, the short circuit and the overturned lantern causing a house fire, and so on, each overdetermining cause plays a distinct and distinctive causal role. The usual notion of overdetermination involves two or more separate and independent causal chains intersecting at a common effect. Because of *Supervenience*, however, that is not the kind of situation we have here. In this sense, this is not a case of genuine causal overdetermination, and *Exclusion* applies in a straightforward way.[86]

Eingeschränktes Überdeterminationsverbot: Die Wirkungen mentaler Ereignisse sind *beinahe nie* genuin überdeterminiert.

Mit diesem eingeschränkten Überdeterminationsverbot ist jedoch kein vollkommen allgemeines Exklusionsargument mehr möglich. Denn auf den oben geschilderten Fall der genuin überdeterminierten Armbewegung lässt sich das übliche Exklusionsprinzip einfach nicht anwenden. Schließlich enthält es eine explizite Ausnahmeklausel für Fälle wie diesen.

Exklusionsargumente, die sich auf das übliche Exklusionsprinzip berufen, unterliegen daher einer nach meiner Einschätzung sehr seltsamen Einschränkung: Sie etablieren, dass mentale Ereignisse *beinahe* keine Wirkungen haben. Wenn aber ein Ereignis von zwei physischen Ereignissen genuin überdeterminiert wird, spricht plötzlich nichts mehr dagegen, auch eine dritte, mentale Ursache anzunehmen. Zumindest steht kein Exklusionsargument zur Verfügung, das dies ausschließen würde.

Natürlich könnte man nun zusätzliche Argumente dafür anführen, dass mentale Ereignisse auch genuin überdeterminierte Ereignisse nicht verursachen können, insofern sie keine nicht genuin überdeterminierten Ereignisse verursachen können. Dies scheint auch plausibel. Der Punkt bleibt aber, dass das übliche Exklusionsprinzip keine streng allgemeine Konklusion erlaubt. Exklusionsargumente müssen dann durch weitere Argumente ergänzt werden, um einen streng allgemeinen Epiphänomenalismus zu etablieren.

[85] Dies hinge dann vom Erfolg der oben angesprochenen Strategien zum Umgang mit genuiner Überdetermination ab.

[86] Kim (2005: 48).

In Fällen von genuiner Überdetermination haben wir es demnach mit zwei *unabhängigen* hinreichenden Ursachen zu tun, während in den Fällen von mentaler Verursachung als Überdetermination, die durch das übliche Exklusionsprinzip ausgeschlossen werden, die mentale Ursache von der hinreichenden physischen Ursache *ontologisch abhängt*: Genuine Überdetermination ist *unabhängige Überdetermination.* Nicht-genuine Überdetermination ist hingegen *abhängige Überdetermination* (siehe Abbildung 3.9 für eine Illustration der Unterscheidung).

Wenn dies der ausschlaggebende Unterschied ist, scheint es recht naheliegend, Exklusionsargumente auf einem Exklusionsprinzip aufzubauen, das sich direkt und ausschließlich gegen abhängige Überdetermination richtet. Mit einem solchen Exklusionsprinzip könnte man zudem auf ein zusätzliches Überdeterminationsverbot verzichten. Es bietet sich also an, das folgende Prinzip ins Zentrum von Exklusionsargumenten zu stellen:

> **Das Exklusionsprinzip der ontologischen Abhängigkeit**: Wenn ein Ereignis u eine (hinreichende) Ursache für ein Ereignis w ist und ein Ereignis h von u ontologisch abhängt, dann ist Ereignis h keine (hinreichende) Ursache für Ereignis w.[87]

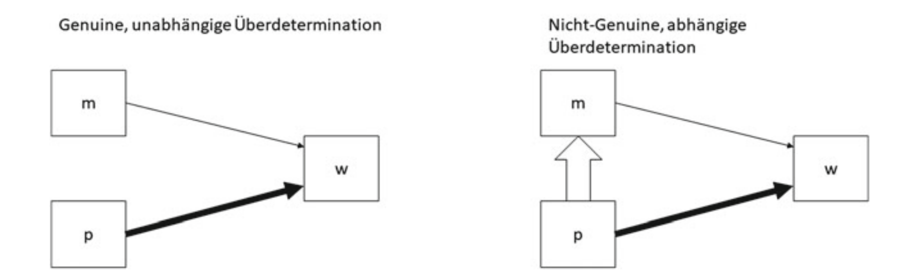
Dieses Prinzip betrifft den Fall der zwei Attentäter ebenso wenig wie das übliche Exklusionsprinzip. Denn im Fall der zwei Attentäter stehen die beiden hinreichenden Ursachen gerade nicht in einer Relation der ontologischen Abhängigkeit. Der Fall der zwei Attentäter ist also kein Gegenbeispiel gegen das Exklusionsprinzip der ontologischen Abhängigkeit.

Auch das Exklusionsprinzip der ontologischen Abhängigkeit kann jedoch für sich genommen nicht die Funktion erfüllen, die das einfache Exklusionsprinzip im einfachen Exklusionsargument übernimmt. Denn im einfachen Exklusionsargument wird an keiner Stelle vorausgesetzt, dass das mentale Ereignis m von der (aufgrund der These der kausalen Geschlossenheit postulierten) hinreichenden physischen Ursache p ontologisch abhängt. Um das Exklusionsprinzip der ontologischen Abhängigkeit zur Anwendung zu bringen, muss man also die These der ontologischen Abhängigkeit hinzuziehen. Insbesondere muss man annehmen, dass die hinreichende physische$_{eng}$ Ursache p von p* zugleich die physische$_{eng}$ Basis von m ist. Man muss also annehmen, dass m von p, der hinreichenden Ursache für p*, ontologisch abhängig ist. Nur unter dieser Voraussetzung ist das Exklusionsprinzip der ontologischen Abhängigkeit auf die relevante Situation anwendbar.

---

[87] Vgl. z. B. List und Menzies (2009).

Auf diese alternative Modifikation des einfachen Exklusionsprinzips soll an dieser Stelle aber lediglich hingewiesen werden. In Teil 4 komme ich an einigen Stellen noch einmal auf den Status des Exklusionsprinzips der ontologischen Abhängigkeit in verschiedenen Theorien der Kausalität zu sprechen. Im Folgenden gehe ich aber zunächst von der verbreiteteren Modifikation aus, die im oben eingeführten üblichen Exklusionsprinzip und dem Überdeterminationsverbot besteht.



Erläuterung: Der dicke Pfeil steht für die Relation der hinreichenden Verursachung. Die dünneren Pfeile stehen für die Relation der (hinreichenden) Verursachung. Der dicke, vertikale Pfeil steht für die Relation der ontologischen Abhängigkeit.
p ist ein physisches$_{eng}$ Ereignis. m ist ein mentales Ereignis. w ist eine Wirkung von m.

Abbildung 3.9 Genuine, unabhängige Überdetermination und nicht-genuine, abhängige Überdetermination

Gegeben diese Unterscheidung schließt also das Exklusionsprinzip aus, dass mentale Verursachung nicht-genuine, abhängige Überdetermination involviert, während das Überdeterminationsverbot ausschließt, dass mentale Verursachung genuine, unabhängige Überdetermination involviert. Gemeinsam schließen die beiden Thesen aus, dass Wirkungen mentaler Ereignisse in irgendeinem Sinne überdeterminiert sind. Auf diese Weise erfüllen sie zusammengenommen die Funktion, die das einfache Exklusionsargument in den bisher dargestellten Exklusionsargumenten übernommen hat.

3.3.10 Die vier Exklusionsargumente – finale Versionen

Die dargestellten Exklusionsargumente sollten also so modifiziert werden, dass sie sich nicht länger auf das einfache Exklusionsargument beziehen, sondern

stattdessen auf die Kombination aus dem üblichen Exklusionsprinzip und dem Überdeterminationsverbot. Es ergeben sich dann die folgenden Versionen der diskutierten Exklusionsargumente:

Das einfache Exklusionsargument

Es sei m ein mentales Ereignis. p und p* seien zwei physische Ereignisse. Unter diesen Annahmen lässt sich das einfache Exklusionsargument wie folgt rekonstruieren:

(1) p ist hinreichende Ursache für p* (motiviert durch die These der kausalen Geschlossenheit).
(2) m ist nicht mit p (oder einem Teil von p) identisch (motiviert durch die These der Nicht-Identität).
(3) Wenn p hinreichende Ursache für p* ist und m nicht mit p (oder einem Teil von p) identisch ist *und p* nicht genuin überdeterminiert ist*, dann ist m nicht Ursache von p* (*motiviert durch das übliche Exklusionsprinzip*).
(4) *p* ist nicht genuin überdeterminiert (motiviert durch das Überdeterminationsverbot).*
(5) Also: m ist nicht Ursache von p* (aus 1, 2, 3, 4).

Erweiterung I: Das Argument aus dem Prinzip der Abwärtsverursachung

Es sei m ein mentales Ereignis. Es sei h ein höherstufiges Ereignis. p und p* seien zwei physische$_{eng}$ Ereignisse. Unter diesen Annahmen lässt sich das Argument aus dem Prinzip der Abwärtsverursachung wie folgt rekonstruieren:

Schritt 1:

(1) h ist ontologisch abhängig von p* (motiviert durch die These der ontologischen Abhängigkeit).
(2) Wenn m Ursache von h ist und h ontologisch abhängig von p* ist, dann ist m Ursache von p* (motiviert durch das Prinzip der Abwärtsverursachung).
(3) Also: Wenn m Ursache von h ist, dann ist m Ursache von p* (aus 1,2).

Schritt 2:

(4) p ist hinreichende Ursache für p* (motiviert durch die These der kausalen Geschlossenheit).
(5) m ist nicht mit p (oder einem Teil von p) identisch (motiviert durch die These der Nicht-Identität).

(6) Wenn p hinreichende Ursache für p* ist und m nicht mit p (oder einem Teil von p) identisch ist *und p* nicht genuin überdeterminiert ist*, dann ist m nicht Ursache von p* (*motiviert durch das übliche Exklusionsprinzip*).
(7) *p* ist nicht genuin überdeterminiert (motiviert durch das Überdeterminationsverbot).*
(8) Also: m ist nicht Ursache von p* (aus 4, 5, 6, 7).
(9) Also: m ist nicht Ursache von h (aus 3, 7).

Erweiterung II: Das Argument aus dem Prinzip der Aufwärtsverursachung
Es sei m ein mentales Ereignis. Es sei h ein höherstufiges Ereignis. p und p* seien zwei $\text{physische}_{\text{eng}}$ Ereignisse. Unter diesen Annahmen lässt sich das Argument aus dem Prinzip der Aufwärtsverursachung wie folgt rekonstruieren:

Schritt 1:

(1) h ist ontologisch abhängig von p* (motiviert durch die These der ontologischen Abhängigkeit).
(2) p ist hinreichende Ursache von p* (motiviert durch die These der kausalen Geschlossenheit).
(3) Wenn p hinreichende Ursache für p* ist und h ontologisch abhängig von p* ist, dann ist p hinreichende Ursache für h.
(4) Also: p ist hinreichende Ursache für h (aus 1, 2, 3).

Schritt 2:

(5) m ist nicht mit p (oder einem Teil von p) identisch (motiviert durch die These der Nicht-Identität).
(6) Wenn p hinreichende Ursache von h ist und m nicht mit p (oder einem Teil von p) identisch ist *und h nicht genuin überdeterminiert ist*, dann ist m nicht Ursache von h (*motiviert durch das übliche Exklusionsprinzip*).
(7) *h ist nicht genuin überdeterminiert (motiviert durch das Überdeterminationsverbot).*
(8) Also: m ist nicht Ursache von h (aus 4, 5, 6, 7).

Erweiterung III: Das Argument aus dem Prinzip der basalen Verursachung
Es sei m ein mentales Ereignis. Es sei w ein beliebiges Ereignis. p und p* seien zwei $\text{physische}_{\text{eng}}$ Ereignisse. Unter diesen Annahmen lässt sich das Argument aus dem Prinzip der basalen Verursachung wie folgt rekonstruieren:

Schritt 1:

(1) m ist ontologisch abhängig von p (motiviert durch die These der ontologischen Abhängigkeit).
(2) Wenn m (hinreichende) Ursache von w ist und m ontologisch abhängig von p ist, dann ist p (hinreichende) Ursache von w (motiviert durch die These der basalen Verursachung).
(3) Also: Wenn m (hinreichende) Ursache von w ist, dann ist p (hinreichende) Ursache von w (aus 1, 2).

Schritt 2:

(4) m ist nicht mit p (oder einem Teil von p) identisch (motiviert durch die These der Nicht-Identität).
(5) Wenn p hinreichende Ursache von w ist und m nicht mit p (oder einem Teil von p) identisch ist *und w nicht genuin überdeterminiert ist*, dann ist m nicht (hinreichende) Ursache von w (*motiviert durch das übliche Exklusionsprinzip*).
(6) *w ist nicht genuin überdeterminiert (motiviert durch das Überdeterminationsverbot).*
(7) Also: Wenn p (hinreichende) Ursache von w ist, dann ist m nicht (hinreichende) Ursache von w (aus 4, 5, 6).
(8) Also: Wenn m (hinreichende) Ursache von w ist, dann ist m nicht (hinreichende) Ursache von w (aus 3, 7).

Wie an diesen Rekonstruktionen deutlich wird, macht der Übergang vom einfachen Exklusionsprinzip zur Kombination aus üblichem Exklusionsprinzip und Überdeterminationsverbot zwei kleine Änderungen in den Argumenten notwendig: Erstens muss die Prämisse, die sich auf ein Exklusionsprinzip beruft, jeweils so modifiziert werden, dass sie Fälle genuiner Überdetermination nicht betrifft. Zweitens muss eine zusätzliche Prämisse eingeführt werden, die besagt, dass die jeweils in Frage stehende vermeintliche Wirkung des mentalen Ereignisses nicht genuin überdeterminiert ist.

Die umliegende Struktur der vier Exklusionsargumente bleibt von diesen Modifikationen jedoch unberührt: Der Übergang vom einfachen Exklusionsprinzip zur Kombination aus üblichem Exklusionsprinzip und Überdeterminationsverbot macht keine Modifikationen der anderen Prämissen der Exklusionsargumente notwendig. Die Prämissen, die sich auf die These der kausalen Geschlossenheit oder auf eines der drei Prinzipien der Verursachung im Stufenmodell der Realität berufen, bleiben unverändert.

Im Folgenden setze ich die gerade genannten finalen Versionen der vier Exklusionsargumente voraus. Wenn ich also vom ‚einfachen Exklusionsargument', vom

‚Argument aus dem Prinzip der Abwärtsverursachung', vom ‚Argument aus dem Prinzip der Aufwärtsverursachung' und vom ‚Argument aus dem Prinzip der basalen Verursachung' spreche, meine ich immer die Versionen, die sich aus der Kombination aus dem üblichen Exklusionsprinzip und dem Überdeterminationsverbot ergeben.

3.3.11 Die vier Exklusionsargumente und der Vorwurf des Epiphänomenalismus

Betrachten wir zum Abschluss noch einmal, wie die vier Exklusionsargumente sich zum Vorwurf des Epiphänomenalismus verhalten. Wie eingangs erwähnt, besteht das Ziel von Exklusionsargumenten im Rahmen des Vorwurfs des Epiphänomenalismus darin, die folgende Prämisse argumentativ zu untermauern:

> **(EP1)** Wenn der nicht-reduktive Physikalismus wahr ist, dann gibt es keine mentale Verursachung.

Um dieses Argumentationsziel zu erreichen, knüpfen Exklusionsargumente an die Kernthesen des nicht-reduktiven Physikalismus an: Mentale Ereignisse sind nicht identisch mit physischen Ereignissen. Jedoch sind mentale Ereignisse (wie auch alle anderen höherstufigen Ereignisse) ontologisch abhängig von physischen Ereignissen. Wenn Exklusionsargumente also nachweisen können, dass diese Kernthesen des nicht-reduktiven Physikalismus dazu führen, dass es keine mentale Verursachung gibt, erreichen sie ihr Argumentationsziel.

Die genannten Kernthesen des nicht-reduktiven Physikalismus implizieren jedoch nicht für sich genommen, dass es keine mentale Verursachung gibt. Die vier Exklusionsargumente, die ich in diesem Kapitel erläutert habe, berufen sich daher auf die eine oder andere Weise auf zusätzliche Annahmen über die kausale Struktur der Welt. Strenggenommen begründen die vier Exklusionsargumente also vorerst den folgenden Zusammenhang zwischen dem nicht-reduktiven Physikalismus und der Existenz mentaler Verursachung:

> **(EP1')** Wenn der nicht-reduktive Physikalismus wahr ist *und eine Reihe von zusätzlichen Thesen über die kausale Struktur der Welt ebenfalls wahr ist*, dann gibt es keine mentale Verursachung.

Zu diesen zusätzlichen Thesen über die kausale Struktur der Welt, die nötig sind, um den nicht-reduktiven Physikalismus in den Epiphänomenalismus zu überführen, zählen die These der kausalen Geschlossenheit, das übliche Exklusionsprinzip, das Überdeterminationsverbot und die drei Prinzipien der Verursachung im Stufenmodell der Realität. Wenn diese Thesen (oder eine geeignete Teilmenge dieser Thesen) also wahr sind, dann führt der nicht-reduktive Physikalismus unweigerlich in den Epiphänomenalismus.

Dies zeigt zugleich, wie nicht-reduktive Physikalist*innen, die an der Existenz mentaler Verursachung festhalten wollen, auf die vier Exklusionsargumente reagieren können: Sie müssen einige der Zusatzannahmen, die nötig sind, um den nicht-reduktiven Physikalismus in den Epiphänomenalismus zu überführen, zurückweisen. Das nächste Kapitel widmet sich nun den Aussichten, die Exklusionsargumente durch Zurückweisung eine der Zusatzannahmen abzuwehren.

3.4 Kausalistische Erwiderungen auf den Vorwurf des Epiphänomenalismus

Die meisten nicht-reduktiven Physikalist*innen lehnen die erste Prämisse des Vorwurfs des Epiphänomenalismus ab. Sie argumentieren dafür, dass es auch dann mentale Verursachung geben kann, wenn der nicht-reduktive Physikalismus wahr ist. Solche kausalistischen Erwiderungen auf den Vorwurf des Epiphänomenalismus stehen nun vor der Aufgabe, die in Abschnitt 3.3. entwickelten Exklusionsargumente zurückzuweisen. In diesem Kapitel behandle ich die Frage, wie Kausalist*innen diese Aufgabe erfüllen können und zu welchen Modellen mentaler Verursachung die verschiedenen Optionen zur Zurückweisung von Exklusionsargumenten führen.

Ich gehe dabei wie folgt vor: In Abschnitt 3.4.1. führe ich eine Unterscheidung zwischen kompatibilistischen und inkompatibilistischen Erwiderungen auf den Vorwurf des Epiphänomenalismus ein: Die kompatibilistische Erwiderung besteht in einer Ablehnung des üblichen Exklusionsprinzips, während die inkompatibilistische Erwiderung in einer Ablehnung der drei Prinzipien der Verursachung im Stufenmodell der Realität besteht. Die darauf folgenden drei Abschnitte beschäftigen sich mit der kompatibilistischen Erwiderung: In Abschnitt 3.4.2. gehe ich im Rahmen dieser Strategie auf den Umgang mit dem Überdeterminationsverbot ein. In Abschnitt 3.4.3. diskutiere und kritisiere ich einige Argumente für das übliche Exklusionsprinzip. Abschnitt 3.4.4. beschäftigt sich mit dem Bild

mentaler Verursachung, das sich aus einer Ablehnung des üblichen Exklusionsprinzips ergibt. In Abschnitt 3.4.5. gehe ich schließlich auf die Aussichten einer inkompatibilistischen Erwiderung auf Exklusionsargumente ein.

3.4.1 Kompatibilismus und Inkompatibilismus

Es gibt zwei vielversprechende Strategien, die vier Exklusionsargumente zurückzuweisen: Die kompatibilistische Strategie und die inkompatibilistische Strategie. In diesem Abschnitt stelle ich knapp die Grundideen der beiden Strategien vor.[88]

Die kompatibilistische Strategie besteht darin, die Kombination aus dem Überdeterminationsverbot und dem üblichen Exklusionsprinzip zurückzuweisen. Die Wirkungen mentaler Ereignisse können dann überdeterminiert sein. Die Strategie besteht also in einer Verteidigung des Modells mentaler Verursachung als (abhängiger) Überdetermination. Diese Strategie kann als ‚Kompatibilismus' bezeichnet werden: Denn im Kern besagt diese Strategie, dass es mit der Existenz mentaler Verursachung kompatibel ist, dass die Wirkungen mentaler Ereignisse bereits hinreichende physische Ursachen haben. Ein und dasselbe Ereignis kann eine hinreichende physische Ursache und darüber hinaus eine weitere mentale Ursache haben.

Das Überdeterminationsverbot und das übliche Exklusionsprinzip kommen in allen vier diskutierten Exklusionsargumenten zur Anwendung. Die kompatibilistische Strategie hat daher den Vorteil, dass sie auf einen Schlag alle Exklusionsargumente zurückweisen kann. Denn wenn die Kombination aus dem Überdeterminationsverbot und dem üblichen Exklusionsprinzip scheitert, dann scheitern alle Exklusionsargumente.

Insofern nur das Überdeterminationsverbot oder das übliche Exklusionsprinzip zurückgewiesen werden, aber die drei Prinzipien der Verursachung im Stufenmodell der Realität akzeptiert werden, hat der Kompatibilismus die Konsequenz, dass *jede* Wirkung eines mentalen Ereignisses überdeterminiert ist. Verantwortlich hierfür sind sowohl die Kombination aus dem Prinzip der Aufwärtsverursachung und der These der kausalen Geschlossenheit als auch das Prinzip der basalen Verursachung: Wie im Argument aus dem Prinzip der Aufwärtsverursachung deutlich wird, sollten die hinreichenden physischen Ursachen der physischen Basen höherstufiger Ereignisse auch als hinreichende Ursachen der zugehörigen höherstufigen Ereignisse gezählt werden. Aus diesem Grund hat nicht nur jedes physische Ereignis eine hinreichende physische Ursache, sondern auch jedes höherstufige

[88] Vgl. für diese Unterscheidung auch Christensen und Baysan (2018: 5–9).

Ereignis. Mentale Ereignisse können also ausschließlich Ereignisse verursachen, die bereits distinkte hinreichende Ursachen haben. Dasselbe folgt auch aus dem Prinzip der basalen Verursachung: Immer, wenn ein mentales Ereignis ein beliebiges anderes Ereignis verursacht, hat auch die physische Basis des mentalen Ereignisses einen Anspruch darauf, eine (hinreichende) Ursache zu sein. Es sind daher alle Wirkungen mentaler Ereignisse durch ein mentales Ereignis und die physische Basis des mentalen Ereignisses überdeterminiert.

Einige Verteidiger*innen des kausalistischen, nicht-reduktiven Physikalismus akzeptieren das Überdeterminationsverbot und das übliche Exklusionsprinzip jedoch (in einem gewissen Sinne[89]): Sie stimmen zu, dass mentale Ereignisse mit ihren physischen Basen um kausalen Einfluss konkurrieren. Jedoch argumentieren sie dafür, dass in allen relevanten Fällen die mentalen Ereignisse als Sieger aus diesem Wettbewerb hervorgehen. So schreibt etwa John Gibbons:

> I think the mental does compete with the physical. I think that determinates compete with determinables, parts compete with wholes, what is realized competes with its realizer, and functional properties compete with the properties that play the roles. I'm not worried about this threat to nonreductive materialism or mental causation because I think that in the relevant cases, the mental properties win.[90]

Diese Strategie zur Erwiderung auf Exklusionsargumente nenne ich ‚Inkompatibilismus'. Inkompatibilist*innen akzeptieren (im Gegensatz zu Kompatibilist*innen) sowohl das Überdeterminationsverbot als auch das übliche Exklusionsprinzip. Sie lehnen jedoch eine geeignete Menge der anderen Zusatzannahmen über die kausale Struktur der Welt ab, die nötig sind, um den nicht-reduktiven Physikalismus in den Epiphänomenalismus zu überführen: Sie lehnen also die These der kausalen Geschlossenheit und/oder die drei Prinzipen der Verursachung im Stufenmodell der Realität ab.

Laut Inkompatibilist*innen sind die Wirkungen mentaler Ereignisse nicht überdeterminiert. Sie haben stattdessen nur mentale Ursachen, aber keine physischen Ursachen. Dies führt auf den ersten Blick zu einer stärkeren kausalen

---

[89] Wie später deutlich werden wird, akzeptieren jedoch auch diese Autor*innen das übliche Exklusionsprinzip, wenn es in Bezug auf *hinreichende* Verursachung formuliert wird. Kompatibilist*innen und Inkompatibilist*innen nehmen lediglich eine unterschiedliche Einstellung zu einer Version des Exklusionsprinzips ein, dass sich auf Verursachung *simpliciter* bezieht. Hieraus folgt zudem eine unterschiedliche Einstellung zu den drei Prinzipien der Verursachung im Stufenmodell der Realität.

[90] Vgl. Gibbons (2006: 82).

Autonomie mentaler Ereignisse: Mentale Ereignisse sind nicht ‚bloß' überdeterminierende Ursachen. Stattdessen haben sie Wirkungen, die nicht zugleich Wirkungen physischer Ereignisse sind.

In Abbildung 3.10 sind die Modelle mentaler Verursachung, die sich aus dem Kompatibilismus und dem Inkompatibilismus ergeben, illustriert.

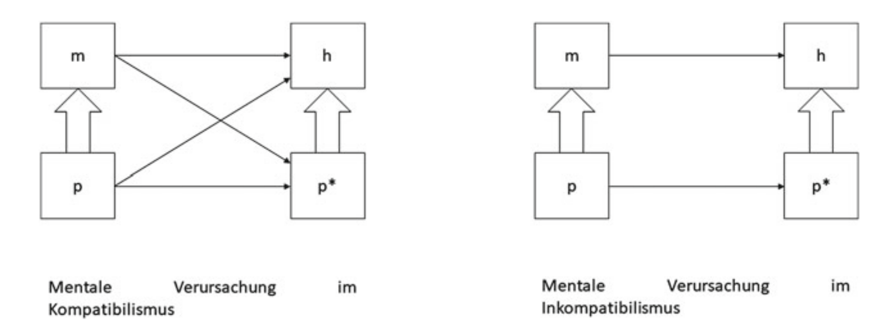


Abbildung 3.10 Kompatibilismus und Inkompatibilismus

3.4.2 Kompatibilismus, das Überdeterminationsverbot und genuine Überdetermination

Das Überdeterminationsverbot besagt, dass die Wirkungen mentaler Ereignisse nicht *genuin* überdeterminiert sind – das heißt, dass sie nicht im selben Sinne überdeterminiert sind, wie der Tod des Opfers im Fall der zwei Attentäter überdeterminiert ist. Es liegen nicht eine hinreichende Ursache und eine *unabhängige* weitere Ursache vor. Die Behauptung wird von Kompatibilist*innen üblicherweise akzeptiert. Und das aus gutem Grund: Denn wenn tatsächlich alle Wirkungen mentaler Ereignisse genuin überdeterminiert wären, wäre das höchst rätselhaft.

Fälle von genuiner Überdetermination involvieren einen großen Zufall oder eine im Hintergrund stehende gemeinsame Ursache. Im Falle der zwei Attentäter können die beiden Attentäter entweder *bloß zufällig* und unabhängig voneinander auf die Idee gekommen sein, einen Anschlag zu verüben. Es müssen weitere Zufälle hinzukommen, damit die beiden Attentäter ihre Anschläge *exakt gleichzeitig* verüben. Die Situation ist dann in der Regel sehr kontrafaktisch instabil: Schon eine kleine Ablenkung einer der beiden Attentäter hätte zum Beispiel dazu geführt, dass er ein wenig später schießt. Alternativ kann eine *gemeinsame*

Ursache dafür vorliegen, dass die beiden Attentäter gleichzeitig schießen: Ein gemeinsamer Auftraggeber wollte beispielsweise sichergehen, dass das Opfer zu genau diesem Zeitpunkt stirbt und hat daher gleich zwei für ihre exakte Pünktlichkeit bekannte Attentäter engagiert. In diesem Falle ist die Situation kontrafaktisch stabiler, aber es stehen menschliche Planung und eine gemeinsame Ursache im Hintergrund.[91] Aus diesen Gründen ist genuine Überdetermination *als allgemeines Modell mentaler Verursachung* abzulehnen: Denn mentale Verursachung ist allgegenwärtig. Dass die mit genuiner Überdetermination verbundene Zufälligkeit oder Planung aber allgegenwärtig ist, scheint überaus unglaubwürdig. Eine Allgegenwärtigkeit von genuiner Überdetermination käme einem gigantischen kosmischen Zufall oder einer rätselhaften kosmischen Verschwörung gleich.[92]

Diese Überlegungen zeigen, dass mentale Verursachung nicht nach dem Modell genuiner Überdetermination verstanden werden sollte. Sie sprechen jedoch nicht dagegen, mentale Verursachung nach dem Modell der abhängigen Verursachung aufzufassen. Denn das Modell der abhängigen Verursachung liefert eine Erklärung für das Zusammentreffen der beiden überdeterminierenden Ursachen gleich mit: Die beiden Ursachen treten regelmäßig zusammen auf, weil eine ontologische Abhängigkeitsbeziehung zwischen ihnen besteht. Die physische Ursache erzwingt die gleichzeitige mentale Ursache, weil die mentale Ursache von der physischen Ursache ontologisch abhängig ist. Es ist also kein bloßer Zufall, dass die beiden Ursachen regelmäßig zusammen auftreten. Und es ist keine Planung nötig, um zu garantieren, dass die beiden Ursachen regelmäßig zusammen auftreten.[93]

Die Beziehung der ontologischen Abhängigkeit, die im Modell mentaler Verursachung als abhängiger Überdetermination zwischen der hinreichenden physischen Ursache und der mentalen Ursache besteht, führt zu weiteren Unterschieden zwischen der Art von Überdetermination, die laut Kompatibilist*innen in Fällen von mentaler Verursachung vorkommt und der genuinen Überdetermination, die im Fall der zwei Attentäter vorliegt. Insbesondere liegen – wie

[91] Vgl. auch Funkhouser (2002) und Moore (2017: 22).

[92] Vgl. auch Sider (2003: 722): „Imagine a paranoiac who thinks that every time someone is shot, there are in fact two causally independent shooters. He is crazy, but why? One reason (not the only one) is that it would be a great coincidence that all these sharp shooters just happen to be at the same places at the same times. This great regularity would need an explanation, and none could be given."

[93] Vgl. z. B. Loewer (2002: 658) und Sider (2003: 722f). Diese Überlegung motiviert zugleich die Annahme, dass die physischen Wirkungen nicht irgendwelche hinreichende physische Ursachen haben, sondern dass die hinreichenden physischen Ursachen der Wirkungen mentaler Ereignisse die physischen Basen der mentalen Ereignisse sind.

Karen Bennett argumentiert[94] – bestimmte kontrafaktische Zusammenhänge, die bei Fällen genuiner Überdetermination vorliegen, bei abhängiger Überdetermination plausiblerweise *nicht* vor.Bei genuiner Überdetermination liegt nämlich in der Regel folgendes kontrafaktisches Muster vor:

> Wenn ein Ereignis w durch ein Ereignis u_1 und ein unabhängiges Ereignis u_2 genuin überdeterminiert ist, dann sind die folgenden beiden Konditionale nicht-leer und wahr:
>
> (a) Wenn u_1 ohne u_2 eingetreten wäre, dann wäre w noch immer eingetreten und
> (b) Wenn u_2 ohne u_1 eingetreten wäre, dann wäre w noch immer eingetreten.

Im Fall der zwei Attentäter wäre das Opfer auch dann gestorben, wenn Attentäter A geschossen hätte, ohne dass Attentäter B geschossen hätte. Und das Opfer wäre auch dann gestorben, wenn Attentäter B geschossen hätte, ohne dass Attentäter A geschossen hätte. Die beiden Konditionale sprechen dabei Möglichkeiten an, die klarerweise sinnvoll sind: Man kann sich leicht vorstellen, dass Attentäter A geschossen hätte, ohne dass Attentäter B schießt.

Dasselbe ist in Fällen von abhängiger Überdetermination nicht der Fall: Denn der modale Zusammenhang zwischen der physischen Basis eines mentalen Ereignisses und dem mentalen Ereignis, der mit der Relation der ontologischen Abhängigkeit einhergeht, hat Konsequenzen für die Auswertung der relevanten kontrafaktischen Konditionale. Wenn ein mentales Ereignis m von einer physischen Basis p ontologisch abhängt, dann *erzwingt* p m. Das heißt: Wenn p eintritt, muss m mit metaphysischer Notwendigkeit ebenfalls eintreten. Dieser Zusammenhang hat die Konsequenz, dass das kontrafaktische Konditional ‚Wenn p ohne m eingetreten wäre, dann wäre w noch immer eingetreten' ein unmögliches Antezedens hat: p kann nicht ohne m eintreten. Dann aber ist das Konditional nach einer verbreiteten Konvention[95] leer wahr.[96] Dies ist eine bedeutende Disanalogie

---

[94] Vgl. für das folgende Bennett (2003).

[95] Vgl. Lewis (1973a: 24–25). Lässt man eine Unterscheidung zwischen falschen und wahren ‚Counterpossibles' (d. h. kontrafaktischen Konditionalen mit unmöglichem Antezedens) zu, muss die Argumentation von Bennett angepasst werden. Möglicherweise wird sie hierdurch auch ernsthaft in Frage gestellt.

[96] Diese Darstellung setzt voraus, dass es sich bei p um eine totale Basis von m handelt. Geht man hingegen von der Kernbasis von m aus, lässt sich unter bestimmten Annahmen argumentieren, dass das kontrafaktische Konditional ‚Wenn p ohne m eingetreten wäre, dann wäre w

zu Fällen genuiner Überdetermination, in denen die entsprechenden Konditionale nicht-leer wahr sind. Gewissermaßen ist die Frage ‚Was wäre passiert, wenn Attentäter A ohne Attentäter B geschossen hätte?' sinnvoll und hat die klare Antwort, dass das Opfer dann noch immer gestorben wäre. Die Frage ‚Was wäre passiert, wenn p ohne m eingetreten wäre?' ist hingegen nicht wirklich sinnvoll und hat keine klare Antwort.

Einige Kompatibilist*innen weisen solchen Disanalogien zwischen genuiner Überdetermination und abhängiger Überdetermination große Bedeutung zu. Bennett meint sogar, dass es die Hauptaufgabe von Kompatibilist*innen sei, Unterschiede zwischen genuiner Überdetermination und abhängiger Überdetermination ausfindig zu machen.[97] Solche Strategien haben nach meiner Auffassung jedoch bestenfalls eine eingeschränkte Überzeugungskraft. Denn sie zeigen zwar, dass die abhängige Überdetermination, die von Kompatibilist*innen für Fälle mentaler Verursachung beansprucht wird, sich von genuiner Überdetermination bedeutend unterscheidet. Sie zeigen daher auch, dass der Kompatibilismus nicht darauf festgelegt ist, dass die Wirkungen mentaler Ereignisse genuin überdeterminiert sind und sie daher nicht auf allgegenwärtige Zufälle oder kosmische Verschwörungen festgelegt sind. Was sie jedoch nicht zeigen, ist, dass das Modell mentaler Verursachung als abhängiger Überdetermination problemlos akzeptiert werden kann.[98] Hierfür müsste vielmehr das übliche Exklusionsprinzip zurückgewiesen werden.

---

noch immer eingetreten' falsch ist. Denn die nächsten Welten, in denen p ohne m eintritt, sind dann solche Welten, in denen p sich nicht in einer geeigneten Umgebung befindet, um m zu necessitieren (z. B. C-Faser Feuern in einem Gehirn, das sich in einer Petrischale befindet). Geht man davon aus, dass die Umgebung, die p benötigt, um m zu necessitieren, zugleich die Umgebung ist, die p benötigt, um w hervorzubringen, würde p in solchen Welten nicht w verursachen. In den nächsten Welten, in denen p ohne m eintritt, würde w also nicht eintreten. Vgl. Bennett (2003: 487–489).

[97] Vgl. Bennett (2003: 474).

[98] Vgl. für eine entsprechende Kritik Morris (2015).

3.4.3 Kompatibilismus und die Begründung des üblichen Exklusionsprinzips

Die Diskussion im letzten Abschnitt hat gezeigt, dass Kompatibilist*innen das Überdeterminationsverbot akzeptieren sollten: Die Wirkungen mentaler Ereignisse sind nicht genuin überdeterminiert. Kompatibilist*innen müssen sich entsprechend auf das übliche Exklusionsprinzip konzentrieren: Sie sollten argumentieren, dass das Modell mentaler Verursachung als abhängiger Überdetermination unproblematisch ist: Es spricht nichts dagegen, dass ein Ereignis eine hinreichende physische Ursache und darüber hinaus eine mentale Ursache hat, solange die mentale Ursache von der hinreichenden physischen Ursache ontologisch abhängt.

Diese Strategie halte ich für durchaus aussichtsreich. Denn letztlich gibt es keine wirklich schlagenden Argumente *für* das übliche Exklusionsprinzip. Um diese Ansicht zu plausibilisieren, gehe ich im Folgenden auf eine Reihe von verbreiteten Argumenten für das übliche Exklusionsprinzip ein.

Dem üblichen Exklusionsprinzip wird von einigen Autor*innen schlicht eine starke intuitive Kraft zugesprochen: Wenn ein Ereignis bereits eine hinreichende Ursache hat, ist einfach unmittelbar klar, dass es nicht eine weitere Ursache haben kann. Zwar mag es hier seltene Ausnahmen geben, die sich in Fällen genuiner Überdetermination zeigen. Diese Ausnahmen sprechen aber nicht gegen die allgemeine Intuition: Mehrere (hinreichende) Ursachen konkurrieren miteinander. Es ist nicht genug Platz für eine hinreichende physische Ursache und eine darüber hinausgehende, zusätzliche, mentale Ursache. Paradigmatisch hierfür zitiere ich Kevin Morris:

> A second salient aspect of my view of the exclusion problem is simply that it has considerable intuitive force. […] Suppose I claim that the raising of my arm has a fully sufficient physical cause, a causal history that involves the contraction of muscles, the firing of neurons, complex physical interactions between my body and the environment, but never steps outside of the physical domain. It is appropriate to find it strange if I go on to assert that my beliefs, desires, and intentions also caused the raising of my arm, if my beliefs, desires, and intentions cannot be identified with some aspect of the physical cause. How can they be causes, one might ask, if there is a complete physical causal story for the raising of your [sic!] arm?[99]

Selbst wenn dem üblichen Exklusionsprinzip eine starke intuitive Kraft zugestanden wird und die Aussage ‚Susis Armbewegung hat eine hinreichende

[99] Morris (2019: 61).

neurologische Ursache und ihr Wunsch verursacht ebenfalls ihre Armbewegung' also kontraintuitiv oder ‚seltsam' ist, kann hieraus meiner Meinung nach keine hinreichende Motivation für die Akzeptanz des Exklusionsprinzips erwachsen.

Denn erstens scheint das übliche Exklusionsprinzip eine substantielle metaphysische Annahme zu sein. Abhängige Überdetermination ist nicht aus begrifflichen Gründen ausgeschlossen. Die Intuition, auf die sich Verteidiger*innen des Exklusionsprinzips beziehen, ist daher keine begriffliche oder semantische Intuition über die Bedeutung der Ausdrücke ‚Verursachung' oder ‚genuine Überdetermination'. Vielmehr handelt es sich um eine Intuition über eine substantielle metaphysische These. Es ist aber fraglich, wieviel Bedeutung man Intuitionen bei der Beantwortung substantieller metaphysischer Fragen zugestehen sollte.

Zweitens führt die Anerkennung einer starken Intuition zugunsten des Exklusionsprinzips und der These, dass Intuitionen eine wichtige Rolle in der metaphysischen Theoriebildung spielen, zunächst lediglich dazu, dass die entsprechende Intuition in einer ‚Gesamtrechnung' der Plausibilität verschiedener theoretischer Optionen berücksichtigt werden sollte. Auch die anderen Annahmen und Prinzipien, die in die Exklusionsargumente einfließen, haben jedoch intuitive Kraft. Wer etwa am Ende die These der mentalen Verursachung oder die These der Nicht-Identität ablehnt, hat ebenso mit kontraintuitiven Konsequenzen der resultierenden metaphysischen Gesamtposition zu kämpfen. Der Verweis auf eine Intuition zugunsten des Exklusionsprinzips etabliert also bestenfalls, dass die Ablehnung des Exklusionsprinzips mit theoretischen Kosten verbunden ist. Ob die zu zahlenden Kosten aber zu hoch sind, bleibt eine offene Frage, die letztlich erst durch eine Abwägung verschiedener Alternativen beantwortet werden kann[100]

Zudem gibt es durchaus auch Überlegungen, die dagegensprechen, dass das übliche Exklusionsprinzip durch eine starke Intuition gedeckt ist. Einige Kompatibilist*innen führen intuitive Gegenbeispiele gegen das übliche Exklusionsprinzip ins Feld. So schreibt etwa Stephen Yablo:

> Remember Archimedes' excited outburst on discovering the principle of displacement in his bath. Assuming that his shouting "Eureka!!" was causally sufficient for his cat's startled flight, nobody would think that this disqualified his (simply) shouting from being causally relevant as well. And it would be incredible to treat Socrates' drinking

[100] Schwitzgebel (2014) argumentiert dafür, dass bei bestimmten metaphysischen Fragestellungen in der Philosophie des Geistes schlichtweg keine Positionen verfügbar sind, die frei von kontraintuitiven oder gar ‚verrückten' Konsequenzen sind.

> the poison as irrelevant to his death, on the ground that his guzzling it was causally sufficient.[101]

Hier wird also ein Fall geschildert, der kein Fall genuiner Überdetermination ist, und dem dennoch mehr als eine (hinreichende) Ursache für ein Ereignis vorliegt: Archimedes Katze erschreckt sich. Hierfür liegen zwei (hinreichende) Ursachen vor: Erstens Archimedes' ‚Heureka'-Schrei und zweitens Archimedes Schrei. Dies sind nach Yablo zwei numerisch distinkte Ereignisse, die jedoch offenbar in einer Relation der ontologischen Abhängigkeit stehen: Der ‚Heureka'-Schrei necessitiert den Schrei, ohne dass der Schrei andersherum den ‚Heureka'-Schrei necessitiert. Yablo meint, dass eine Konkurrenzintuition hier gar nicht erst aufkommt: *Natürlich* stehe der ‚Heureka'-Schrei nicht in einer Konkurrenz um kausalen Einfluss mit dem Schrei.

Wer diese Intuition teilt, muss das übliche Exklusionsprinzip als allgemeines Prinzip ablehnen: Wenn zwei Ereignisse in bestimmten, sehr engen metaphysischen Abhängigkeitsbeziehungen stehen, konkurrieren sie intuitiv gerade *nicht* um kausalen Einfluss. Es gibt daher Fälle, in denen zwei numerisch distinkte Ereignisse hinreichende Ursachen für eine Wirkung bilden, ohne dass diese Wirkung genuin überdeterminiert ist.

Natürlich gibt es einige Kommentator*innen, die Yablos Intuition nicht teilen und argumentieren, dass es sehr wohl eine Konkurrenz um kausalen Einfluss zwischen Archimedes' ‚Heureka'-Schrei und seinem Schrei gibt.[102] Was Yablos Berufung auf Intuition aber zeigt, ist meines Erachtens, dass die Intuitionen, die mit dem üblichen Exklusionsprinzip verbunden sind, keinesfalls eindeutig sind. Dies allein macht die Berufung auf Intuition von beiden Seiten letztlich problematisch: Weder die Befürworter*innen noch die Kritiker*innen des Exklusionsprinzips können sich auf eine klare und eindeutige Intuition berufen. Wie so oft kann die Berufung auf Intuition hier keine allgemein akzeptierte Entscheidung herbeiführen.

Eine Berufung auf Intuition zur Verteidigung des üblichen Exklusionsprinzips ist aber auch gar nicht notwendig, wenn starke *Argumente* für das übliche Exklusionsprinzip angeführt werden können. Zwei solcher Argumente möchte ich im Folgenden noch diskutieren: Erstens eine Berufung auf ein allgemeines

---

[101] Yablo (1992: 272). Yablo will hiermit die Intuition hervorrufen, dass Determinablen nicht mit ihren Determinaten um kausalen Einfluss konkurrieren. Explizit als Gegenbeispiel führen Árnadóttir und Crane (2013: 258) zudem einen Fall an, indem wir sowohl gewillt sind, ein ganzes Objekt als (hinreichende) Ursache für ein Ereignis zu zählen, als auch gewillt sind, einen Teil des Objekts als (hinreichende) Ursache zu zählen.

[102] Vgl. z. B. Walter (2007: 231–234).

Prinzip der Einfachheit und zweitens eine Berufung auf eine Auffassung von Verursachung als das Leisten kausaler Arbeit.

Das Prinzip der Einfachheit besagt grob gesprochen folgendes: Man soll nicht mehr Entitäten postulieren, als unbedingt nötig. Einfache ontologische Theorien, die mit weniger Entitäten auskommen, sind weniger einfachen Theorien *ceteris paribus* vorzuziehen. Wie lässt sich dieses Prinzip der Einfachheit auf die Situation abhängiger Überdetermination beziehen? Kim schreibt hierzu:

> First, there is the good old principle of simplicity: we can make do with P as a cause of P*, so why bother with M?[103]

Kim wendet das Prinzip der Einfachheit hier direkt auf eine vermeintliche Situation der abhängigen Überdetermination an: Die hinreichende physische Ursache p reicht bereits aus, um ohne Rest zu erklären, warum p* eintritt. Es ist daher nicht nötig, anzunehmen, dass auch das mentale Ereignis m, das von p ontologisch abhängt, p* verursacht. Das Prinzip der Einfachheit diktiert daher, auf diese unnötige Annahme zu verzichten: Wenn p* also bereits eine hinreichende Ursache hat, sollten wir keine weiteren Verursachungsrelationen mehr postulieren.[104]

Liefert diese Überlegung ein überzeugendes Argument für das übliche Exklusionsprinzip? Ich sehe eine Reihe von Problemen:

Erstens macht es stutzig, dass die Argumentation auf den ersten Blick auch auf Fälle von genuiner Überdetermination anwendbar zu sein scheint: Wenn schon der Schuss des ersten Attentäters eine hinreichende Ursache für den Tod des Opfers ist, wieso sollten wir dann davon ausgehen, dass auch der Schuss des zweiten Attentäters eine Ursache ist? Schließlich benötigen wir diese Annahme nicht, um zu erklären, dass das Opfer gestorben ist. Sollten wir diese zweite Ursache also aus Einfachheitsgründen wegkürzen? Das ist nicht

---

[103] Kim (1993b: 207).

[104] Vgl. auch Carey (2011: 254) und Engelhardt (2015: 207–208). Tatsächlich muss es in dieser Überlegung um die Anzahl postulierter *Verursachungsrelationen* gehen, und nicht etwa um die Anzahl postulierter *Ereignisse*. Die vorliegende Anwendung des Einfachheitsprinzips muss also von zwei anderen möglichen Anwendungen in anderen Kontexten unterschieden werden: Erstens wird das Prinzip häufig auf die Annahme angewandt, *dass m existiert*. Gegeben, dass m keine Wirkungen hat, solle man aus Gründen der Einfachheit auf diese Annahme verzichten. Zweitens wird das Prinzip häufig auf die Annahme angewandt, *dass m nicht identisch mit p ist*. Gegeben, dass m dieselben Wirkungen wie p hat, solle man aus Gründen der Einfachheit auf diese Annahme verzichten. Beide Anwendungen vermögen nicht, das Exklusionsprinzip zu begründen. Denn im üblichen Exklusionsprinzip geht es um die Frage, ob m eine Ursache für p* sein kann, *gegeben dass m existiert und nicht identisch mit p ist*.

plausibel. Kim schuldet uns daher eine Erklärung, warum das Prinzip der Einfachheit diktiert, abhängige Überdeterminierer wegzukürzen, während unabhängige Überdeterminierer unberührt bleiben.

Eine mögliche Erklärung ist folgende: In Fällen von genuiner Überdetermination ist es möglich, die beiden Ursachen zu trennen und ihr Wirken in Isolation zu betrachten. Wir wissen aus unabhängigen Quellen, dass Schüsse wie der des zweiten Attentäters hinreichend für Tode wie dem des Opfers sind. Solche Schüsse kommen in anderen Kontexten auch alleine vor und verursachen Tode. Das scheint ein wichtiger Grund zu sein, warum in Fällen von genuiner Überdetermination beide Ereignisse als (hinreichende) Ursachen gezählt werden sollten. Dieser Grund steht jedoch in Fällen von abhängiger Überdetermination nicht zur Verfügung: Physische Basen mentaler Ereignisse kommen notwendigerweise nicht ohne die zugehörigen mentalen Ereignisse vor. Und mentale Ereignisse kommen nie ohne eine physische Basis vor. Die Idee ist also, dass mentale Ereignisse in Fällen von abhängiger Überdetermination *immer* überdeterminierend sind, während die hinreichenden Ursachen in Fällen von genuiner Überdetermination auch als nicht-überdeterminierende Ursachen auftreten.[105]

Das Einfachheitsprinzip richtet sich, so ließe sich argumentieren, nur auf Ursachen, die *immer* überdeterminierend sind und lässt Fälle von genuiner Überdetermination deshalb unberührt. Wir benötigen m nicht, um zu erklären, dass p* eintritt, weil *jedes Mal*, wenn ein p*-ähnliches Ereignis auf ein m-ähnliches Ereignis folgt, auch eine hinreichende physische Ursache p vorliegt. Es ist dieser Umstand, der es erlaubt, m als Ursache von p* wegzukürzen.

Einfachheitsüberlegungen entwickeln daher erst Überzeugungskraft, wenn sie auf das Gesamtbild angewendet werden, das vom Modell der mentalen Verursachung als abhängiger Überdetermination impliziert wird. Kevin Morris schreibt hierzu:

> An outlook that involves occurrences that are always "overdetermining causes" if they are causes at all plainly offends against parsimony, and other things being equal an outlook that does not posit occurrences that, while causes, are always causes of events that also have sufficient but distinct causes is to be preferred.[106]

Hier wird das Prinzip der Einfachheit auf die These angewendet, dass mentale Ereignisse *ausschließlich* überdeterminierte Wirkungen haben. Wenn sich eine solche These vermeiden lässt, sollte sie vermieden werden. Denn eine Theorie, die ständig zwei Verursachungsrelationen postuliert, wo auch eine reichen würde,

[105] Vgl. Morris (2019: 57–59).

[106] Morris (2019: 59).

ist *ceteris paribus* weniger einfach als eine Theorie, die auf eine solche Annahme verzichtet.[107]

Doch auch dieses Argument aus der Einfachheit kann nicht völlig überzeugen. Denn die Auswertung der Einfachheit von Theorien ist sehr umstritten:

Es ist zum einen nicht klar, an welchen explanatorischen Aufgaben die Einfachheit einer Theorie gemessen wird: Welche Dinge sollen eigentlich durch die Theorie erklärt werden? Und wozu genau ist die Annahme einer weiteren Verursachungsrelation zwischen m und p* überflüssig, wenn p* bereits eine hinreichende Ursache hat? Selbst wenn man zugesteht, dass man die Verursachungsrelation zwischen m und p* nicht annehmen muss, um das Auftreten von p* vollständig zu erklären, folgt hieraus nicht, dass die Annahme dieser Verursachungsrelation *schlechthin* explanatorisch irrelevant ist. Denn diese Annahme hat durchaus eine theoretische Funktion: Sie ermöglicht beispielsweise eine kausale Handlungstheorie.

Zum anderen ist unklar, wie genau die Postulierung bestimmter Entitäten in die Bewertung der Einfachheit einer Theorie einfließt – selbst wenn das explanatorische Ziel feststeht. Karen Bennett argumentiert etwa dafür, dass die Annahme zusätzlicher *nicht-fundamentaler, höherstufiger* Entitäten nicht gegen die Einfachheit einer Theorie ins Feld geführt werden sollte.[108] Denn die Existenz nicht-fundamentaler Entitäten ergibt sich einfach aus der Existenz fundamentaler Entitäten. Es handelt sich nicht wirklich um zusätzliche, neue Annahmen, die nicht bereits in Annahmen über die Existenz fundamentaler Entitäten implizit drinsteckte. Diese Überlegung hat Konsequenzen für die Postulierung der Verursachungsrelation zwischen m und p*: Denn die Annahme, dass m p* verursacht, spricht nur dann gegen die Einfachheit der resultierenden Theorie mentaler Verursachung, wenn diese Relation selbst fundamental ist und sich nicht aus anderen, fundamentalen Sachverhalten ergibt. Dies müssen kausalistische nicht-reduktive Physikalist*innen aber gar nicht behaupten.

[107] Man sollte jedoch darauf hinweisen, dass nicht-reduktive Physikalist*innen das übliche Exklusionsprinzip ablehnen können, ohne das Modell abhängiger Überdetermination als *universelles* Modell mentaler Verursachung zu akzeptieren. Eine Ablehnung des Exklusionsprinzips, demzufolge abhängige Überdetermination nie vorkommt, legt niemanden darauf fest, dass jeder Fall von mentaler Verursachung ein Fall von abhängiger Überdetermination ist. Nicht-reduktive Physikalist*innen müssen allerdings zusätzlich zum Exklusionsprinzip auch das Prinzip der Aufwärtsverursachung und das Prinzip der basalen Verursachung ablehnen, um die Konklusion zu vermeiden, dass jeder Fall von mentaler Verursachung ein Fall von abhängiger Überdetermination ist.

[108] Vgl. Bennett (2017: 220–225).

Eine andere Begründung des üblichen Exklusionsprinzips beruft sich auf Überlegungen zur Metaphysik der Verursachung. Kim motiviert das Exklusionsprinzip an verschiedenen Orten[109] unter Berufung auf den Begriff der *kausalen Arbeit*. Paradigmatisch hierfür ist seine folgende Frage:

> Given that p has a physical cause p*, what causal work is left for m to contribute?[110]

Was ist hier mit ‚kausaler Arbeit' gemeint?[111] Der Begriff ist leider sehr dunkel. An dieser Stelle genügt es, einige Metaphern anzuführen, die mit der Idee von kausaler Arbeit üblicherweise verknüpft werden: Wenn Ursachen kausale Arbeit leisten, dann *übertragen* sie irgendetwas auf ihre Wirkungen. Sie *produzieren* ihre Wirkungen und *bringen sie hervor*. Verursachung involviert dann – um die vielleicht dunkelste, aber auch verbreitetste Metapher in dieser Gruppe zu bemühen – ‚*causal oomph*'.

Auch ist nicht ganz klar, wie die Auffassung von Verursachung als das Leisten kausaler Arbeit das übliche Exklusionsprinzip eigentlich motivieren soll. Sven Walter rekonstruiert eine mögliche Motivation wie folgt:

> Sobald die ‚kausale Arbeit', die für das Auftreten einer Wirkung erforderlich ist, einmal von einer Eigenschaft verrichtet wurde, muss sie nicht nur nicht mehr von einer anderen Eigenschaft verrichtet werden, sie *kann* dann nicht mehr von einer anderen Eigenschaft verrichtet werden. ‚Kausale Arbeit' ist ganz einfach nichts, was in irgendeinem Sinne ‚doppelt' getan werden könnte – ob die fragliche Überdetermination systematisch oder koinzidentell ist, spielt dabei keine Rolle.[112]

---

[109] Vgl. z. B. Kim (1990: 45), Kim (1993b: 207), Kim (1998: 37, 45, 53), Kim (2002b: 675), Kim (2005: 45) und Kim (2007: 236).

[110] Kim (1998: 37).

[111] Was versteht Kim unter ‚kausaler Arbeit'? An einigen Stellen bringt er diesen Begriff mit Phil Dowes (2000) Erhaltungsgrößentheorie der Kausalität zusammen: Das Leisten kausaler Arbeit involviert demzufolge das Erhalten oder Übertragen einer physikalischen Erhaltungsgröße. Auf diese Idee gehe ich in Abschnitt 4.4. noch genauer ein. An anderen Stellen beruft er sich auf G.E.M. Anscombes (1971) klassische Arbeit zu Kausalität. Zudem verweist er gern auf die Metaphern, die ich im Fließtext anführe.

[112] Walter (2008b: 113). Ganz ähnlich schreibt Block (2003: 135): „Perhaps the thought would be that if there is a sufficient physical cause and a sufficient mental cause, then each does 100 % of the work, and so more work is being done than there is to do". Vgl. auch Sider (2003: 721): „Here is a picture. Causation is a kind of fluid divided among the potential causes of an effect. If one potential cause acts to produce an effect, that fluid is *used up*, and no other potential causes can act". Alle drei Autoren lehnen diese Auffassung von Kausalität jedoch ab.

Hier ist die Idee also, dass es ein gewisses Maß an kausaler Arbeit gibt, das für die Hervorbringung einer Wirkung nötig ist. Zugleich ist es unmöglich, dass mehr kausale Arbeit verrichtet wird, als nötig ist, um eine Wirkung hervorzubringen. Da eine hinreichende Ursache bereits die volle Arbeit verrichtet, kann es keine zusätzliche Wirkung geben, die ebenfalls Arbeit verrichtet.

Diese Argumentation hat erstens das Problem, dass sie nicht das übliche sondern das einfache Exklusionsprinzip rechtfertigt. Wenn dies die richtige Auffassung von Verursachung und kausaler Arbeit ist, dann kann es allgemein keine Fälle von Überdetermination geben. Denn im Fall der zwei Attentäter würde auf den ersten Blick ebenfalls ‚doppelte' kausale Arbeit geleistet. Wenn dies aber nicht vorkommen kann, dann ist auch der Fall der zwei Attentäter nicht möglich. Entsprechend würde es für eine Zurückweisung des Exklusionsprinzips reichen, für die prinzipielle Möglichkeit genuiner Überdetermination zu argumentieren.

Zweitens beruht die Argumentation auf dem dunklen Begriff der kausalen Arbeit. Wenn dieser Begriff zudem präzisiert wird, passiert dies oft auf der Grundlage einer Theorie der Kausalität, die bereits für sich genommen gegen die Möglichkeit mentaler Verursachung im nicht-reduktiven Physikalismus spricht. Kim verweist an entsprechenden Stellen auf Phil Dowes Erhaltungsgrößentheorie der Kausalität, der zufolge Verursachung in dem Übertragen oder Erhalten von physikalischen Erhaltungsgrößen besteht.[113] Diese Theorie impliziert jedoch schon für sich genommen bereits einen umfassenden Reduktionismus in Bezug auf alle kausal effektiven Ereignisse.[114] Dies ist dialektisch für Verteidiger*innen von Exklusionsargumenten sehr ungünstig: Denn wenn das Exklusionsprinzip eine Theorie der Kausalität voraussetzt, die für sich genommen garantiert, dass es im nicht-reduktiven Physikalismus keine mentale Verursachung gibt, scheint das Exklusionsargument überflüssig zu werden.[115]

Andrew Russo schlägt eine andere Begründung des Exklusionsprinzips vor, die ihren Ausgang von dem Begriff der kausalen Arbeit nimmt:

---

[113] Vgl. Dowe (2000).

[114] Vgl. Dowe (2009: 224): „In any case, to suppose that the conserved quantity theory will deal with causation in other branches of science also requires commitment to a fairly thorough going reductionism, since clearly there is nothing in economics or psychology that could pass for a conservation law".

[115] Vgl. Walter (2008: 112–117) und Raatikainen (2018: 41): „In other words, if the exclusion argument in reality presupposes a specific theory of causation, and that theory in turn presupposes a strong and all-encompassing reductive view, it seems to beg the question to appeal to the exclusion argument in defense of a reductive view. That is to say, the exclusion argument seems to become quite pointless".

> The idea that causation requires a "flow" or "transfer" of something between cause and effect suggests a picture of overdetermined bodily effects as receiving something from their neurophysiological cause and an *additional something* from their mental cause. It is this additional something which creates the problem: in order to explain why an overdetermined bodily effect has as much of that something as it has, we would have to make an essential appeal to an irreducible mental event; citing its neurophysiological antecedents would be explanatorily insufficient. Thus, systematic overdetermination involving productive causation would result in the systematic failure of a certain kind of physical explanation.[116]

Hier wird nicht vorausgesetzt, dass kausale Arbeit prinzipiell nicht im Überfluss geleistet werden kann. Stattdessen ist die Idee, dass ein Überfluss an kausaler Arbeit eine Erklärung benötigen würde. Wenn die überschüssige kausale Arbeit von einer nicht-physischen Ursache geleistet wird, ließe sich keine physische Erklärung angeben. Es wäre nötig, auf eine nicht-physische Ursache zu verweisen, um die Wirkung vollständig zu erklären. Dies wiederum würde gegen die prinzipielle explanatorische Vollständigkeit der Physik und damit letztlich gegen die These der kausalen Geschlossenheit sprechen.[117]

Diese Begründung spricht jedoch nicht für ein Exklusionsprinzip, das neutral gegenüber der Unterscheidung zwischen physischen und nicht-physischen Ereignissen ist. Denn solange die überschüssige kausale Arbeit von einer überdeterminierenden *physischen* Ursache geleistet wird, ist auch eine physische Erklärung möglich und die explanatorische Vollständigkeit der Physik gerät nicht in Gefahr.

Tatsächlich legt Russos Begründung nahe, dass nicht-physische Ereignisse im Allgemeinen keine kausale Arbeit für das Eintreten physischer Ereignisse leisten können – ganz gleich, ob ihre Wirkungen überdeterminiert sind oder nicht. Denn jede ‚Injektion' kausaler Arbeit vom Mentalen in das Physische würde zu einer Verletzung der explanatorischen Vollständigkeit der Physik und der These der kausalen Geschlossenheit führen. Zugleich würde sich ein so begründetes Exklusionsprinzip zunächst nur auf mental-physische Abwärtsverursachung

[116] Russo (2015: 2188).

[117] Diese Begründung spricht, soweit ich sehe, nicht direkt gegen die prinzipielle Möglichkeit genuiner Überdetermination: Solange der Tod des Opfers im Fall der zwei Attentäter im relevanten Sinne derselbe wäre, auch wenn weniger kausale Arbeit geleistet worden wäre, kann der Tod des Opfers genuin überdeterminiert sein. Die Menge an geleisteter kausaler Arbeit darf aber nicht essentiell für den Tod des Opfers sein, da anderenfalls Bunzls (1979) Strategie zur Vermeidung genuiner Überdetermination greifen würde: *Dieses maximal spezifische* Ereignis (mit *dieser* Menge an geleisteter kausaler Arbeit) wird nur von beiden Ursachen gemeinsam hervorgebracht.

und nicht auf mental-höherstufige Abwärtsverursachung oder mentale Intralevel-Verursachung anwenden lassen. Denn wenn mentale Ereignisse kausale Arbeit für andere mentale oder höherstufige Ereignisse leisten, spricht dies nicht direkt gegen die explanatorische Vollständigkeit der Physik.

Dies sorgt für zwei Probleme eines so begründeten Exklusionsprinzips im Kontext von Exklusionsargumenten: Erstens könnte ein so begründetes, nicht-neutrales Exklusionsprinzip allenfalls im einfachen Exklusionsargument und im Argument aus dem Prinzip der Abwärtsverursachung zur Anwendung kommen. Denn nur hier wird das Exklusionsprinzip ausschließlich auf physische Wirkungen angewendet. Im Argument aus dem Prinzip der Aufwärtsverursachung und im Argument aus dem Prinzip der basalen Verursachung wird hingegen ein Exklusionsprinzip benötigt, das sich auch auf nicht-physische Wirkungen anwenden lässt.

Zweitens macht auch diese Begründung das einfache Exklusionsargument überflüssig. Denn sie beruht darauf, dass Verursachung auf geeignete Weise mit kausaler Arbeit verknüpft ist und dass kausale Arbeit für physische Ereignisse nicht durch nicht-physische Ereignisse geleistet werden kann. Wenn das aber stimmt, braucht es kein Exklusionsargument mehr, um zu zeigen, dass mental-physische Verursachung im nicht-reduktiven Physikalismus nicht möglich ist.[118]

Letztlich, so meine ich, kann keine der verbreiteten Begründungen des üblichen Exklusionsprinzips vollständig überzeugen. Die Ablehnung des üblichen Exklusionsprinzips stellt daher durchaus eine verfolgenswerte Option dar.

3.4.4 Kompatibilismus, das Ausmaß der Überdetermination und die Redundanz mentaler Ursachen

Welche Auffassung von mentaler Verursachung folgt aber aus einer Ablehnung des üblichen Exklusionsprinzips? Und handelt es sich dabei um eine insgesamt zufriedenstellende Auffassung von mentaler Verursachung?

Soweit die anderen Thesen über die kausale Struktur der Welt, die in die vier Exklusionsargumente einfließen, akzeptiert werden, hat der Kompatibilismus die Konsequenz, dass *jede* Wirkung eines mentalen Ereignisses auch eine hinreichende physische Ursache hat. Mentale Verursachung ist demnach *immer* überdeterminierende Verursachung: Zwar haben mentale Ereignisse Wirkungen. Sie haben jedoch in einem gewissen Sinne keine *eigenständigen* Wirkungen.

[118] Vgl. wiederum Walter (2008b: 112–117) und Raatikainen (2018: 40–41).

Denn ihre Wirkungen sind immer auch Wirkungen gleichzeitig stattfindender physischer Ereignisse.

Dies mag unbefriedigend wirken. Es bleibt das ungute Gefühl, dass mentale Ursachen *redundant* sind. Sie verursachen nur, was auch ohne sie durch physische Ereignisse verursacht werden würde. Und wenn es gar keine mentalen Ereignisse gäbe, würde das letztlich keinen Unterschied für den Verlauf der Welt machen. Denn dann gäbe es noch immer die physischen Ursachen, die für denselben Verlauf der Welt sorgen würden.

Diesen Bedenken kann man auf zweierlei Weise entgegenkommen: Erstens kann man das Ausmaß an Überdetermination im Kompatibilismus abschwächen, indem man zusätzlich zum üblichen Exklusionsprinzip auch die drei Prinzipien der Verursachung im Stufenmodell der Realität ablehnt. Insbesondere muss hierfür das Prinzip der Aufwärtsverursachung und das Prinzip der basalen Verursachung abgelehnt werden. Wenn diese Prinzipien abgelehnt werden, ist es möglich, dass die höherstufigen Wirkungen mentaler Ereignisse keine (hinreichenden) physischen Ursachen haben. Dann ist nicht jeder Fall von mentaler Verursachung ein Fall von Überdetermination. Stattdessen sind nur Fälle von mental-fundamentaler Abwärtsverursachung immer Fälle von Überdetermination.

Zweitens kann man dafür argumentieren, dass der Umstand, dass mentale Ereignisse von ihren physischen Basen ontologisch abhängen, dazu führt, dass man eben nicht sinnvoll behaupten kann, dass abhängig überdeterminierende mentale Ereignisse in einem problematischen Sinne redundant sind. Die kontrafaktischen Überlegungen, die die Behauptung der Redundanz auf den ersten Blick plausibel machen, sind nämlich nur im Falle von genuiner Überdetermination auf eine nicht-leere Weise wahr: Natürlich würde der Schuss von Attentäter A auch in Abwesenheit von Attentäter B zum Tod des Opfers führen (und andersherum). Aber es ist nicht auf dieselbe Weise wahr, dass die physische Basis von Susis Wunsch auch in Abwesenheit von Susis Wunsch dazu führen würde, dass Susis ihren Arm hebt. Denn während es eine völlig problemfreie und mögliche Situation gibt, in der nur ein Attentäter schießt, gibt es keine mögliche Situation, in der die Basis von Susis Wunsch ohne ihren Wunsch vorkommt. Auch an dieser Stelle sind also die Überlegungen zur Disanalogie zwischen genuiner Überdetermination und abhängiger Überdetermination relevant.

3.4.5 Inkompatibilismus

Dennoch ist es eine interessante Frage, ob kausalistische nicht-reduktive Physikalist*innen das übliche Exklusionsprinzip prinzipiell auch akzeptieren können.

Ist die Existenz mentaler Verursachung im nicht-reduktiven Physikalismus also damit vereinbar, dass die Wirkungen mentaler Ereignisse *nicht* überdeterminiert sind?

Inkompatibilist*innen meinen, diese Frage bejahen zu können. Die Wirkungen mentaler Ereignisse haben ausschließlich mentale Ursachen, aber keine physischen Ursachen.[119] Und wie meine Darstellung von Exklusionsargumenten gezeigt hat, ist dies durchaus eine theoretische Möglichkeit. Denn die These der kausalen Geschlossenheit etabliert zunächst nur, dass $\text{physische}_{\text{eng}}$ Ereignisse hinreichende $\text{physische}_{\text{eng}}$ Ursachen haben. Zieht man nun das Überdeterminationsverbot und das übliche Exklusionsprinzip hinzu, kann durch das einfache Exklusionsargument gezeigt werden, dass mentale Ereignisse keine $\text{physischen}_{\text{eng}}$ Wirkungen haben. Aber dass mentale Ereignisse keine $\text{physischen}_{\text{eng}}$ Wirkungen haben, bedeutet nicht, dass sie gar keine Wirkungen haben. Mentale Ereignisse können insbesondere im Einklang mit der Konklusion des einfachen Exklusionsarguments behaviorale und mentale Wirkungen haben. Und diese Wirkungen mentaler Ereignisse müssen keine hinreichenden physischen Ursachen haben. Zumindest schreibt die These der kausalen Geschlossenheit für sich genommen nicht vor, dass behaviorale und mentale Ereignisse hinreichende physische Ursachen haben.

Erst die Erweiterungen des einfachen Exklusionsarguments schließen diese Lösung aus. Diese berufen sich aber neben dem Überdeterminationsverbot und dem einfachen Exklusionsprinzip auch auf die drei Prinzipien der Verursachung im Stufenmodell der Realität. Es ist daher auf den ersten Blick eine theoretische Option, diese Erweiterungen nicht durch eine Ablehnung des üblichen Exklusionsprinzips, sondern durch eine Ablehnung der drei Prinzipien der Verursachung im Stufenmodell der Realität zurückzuweisen. Wenn dies gelingt, können nicht-reduktive Physikalist*innen das übliche Exklusionsprinzip akzeptieren und zugleich an der Existenz mentaler Verursachung festhalten.

Inkompatibilist*innen müssen hierfür jedoch alle drei Prinzipien der Verursachung im Stufenmodell der Realität ablehnen. Denn jedes der drei Prinzipien führt zu einem eigenständigen Exklusionsargument, das jeweils etabliert, dass mentale Ereignisse keinerlei Wirkungen haben. Es muss also jedes der drei Argumente zurückgewiesen werden. Und das bedeutet, alle drei Prinzipien der Verursachung im Stufenmodell der Realität zurückzuweisen.

Die Aussichten dieser Strategie können besser im Rahmen von konkreteren Annahmen über die Natur der Verursachungsrelation diskutiert werden.

[119] Vgl. Gibbons (2006), List und Menzies (2009), Zhong (2014) und Pernu (2016) für verschiedene inkompatibilistische Ansätze.

Später wird sich zeigen, dass der Inkompatibilismus nicht *jede* Version von Exklusionsprinzipien akzeptieren kann. Denn die Einsicht, dass höherstufige Ereignisse *hinreichende* physische Ursachen haben, ist auch im Inkompatibilismus kaum zu vermeiden. Inkompatibilist*innen können jedoch argumentieren, dass hieraus nicht folgt, dass höherstufige Ereignisse physische Ursachen *simpliciter* haben. Und dies erlaubt ihnen, Versionen des Exklusionsprinzips zu akzeptieren, die unter Bezug auf Verursachung simpliciter formuliert sind. Dabei setzt der Inkompatibilismus voraus, dass Ursachen proportional zu ihren Wirkungen sind: Ursachen enthalten weder zu viel überflüssiges Detail noch zu wenig notwendiges Material, um ihre Wirkungen hervorzubringen. Unter diesen Voraussetzungen können bestimmte Versionen der drei Prinzipien der Verursachung zurückgewiesen werden. Die genauere Diskussion dieser Punkte verschiebe ich jedoch auf Teil 4 dieser Arbeit.[120]

3.5 Die Existenz mentaler Verursachung

Der Vorwurf des Epiphänomenalismus wird erst deshalb zu einem Problem für den nicht-reduktiven Physikalismus, weil die Existenz mentaler Verursachung als unhintergehbares Faktum vorausgesetzt wird. Der Epiphänomenalismus – der dieses scheinbar unhintergehbare Faktum abstreitet – hat in erster Linie die Rolle eines Schreckgespenstes. Wenn der nicht-reduktive Physikalismus daher den Epiphänomenalismus zur Konsequenz hat, gilt er als inakzeptabel. Paradigmatisch für diese Ansicht sind Jerry Fodors Kommentare zum Epiphänomenalismus:

> An outbreak of epiphobia (epiphobia is the fear that one is turning into an epiphenomenalist) appears to have much of the philosophy of mind community in its grip. [Epiphobics] fear that the very successes of a physicalistic (and/or a computational) psychology will entail the causal inertness of the mental. Fearing this makes them unhappy.[121]
>
> [I]f it isn't literally true that my wanting is causally responsible for my reaching, and my itching is causally responsible for my scratching, and my believing is causally responsible for my saying ... if none of that is literally true, then practically everything I believe about anything is false and it's the end of the world.[122]

---

[120] Siehe insbesondere 4.1.3., 4.2.3., 4.3.12.–4.3.14. und 4.4.10.

[121] Fodor (1989: 59).

[122] Fodor (1989: 77). Ähnlich auch Ewing (1953: 127): „That epiphenomenalism is false is assumed in all practical life“ oder Bontly (2005: 331): „[The existence of mental causation]

Die ‚Epiphobie', die laut Fodor auch viele (nicht-reduktiv) *physikalistischen* Philosoph*innen befallen hat, ergibt sich dabei zunächst aus den Exklusionsargumenten. Die überspitzte Rede vom ‚Ende der Welt', das mit einer Akzeptanz des Epiphänomenalismus einhergehe, verweist auf die Zentralität, die die Existenz mentaler Verursachung innerhalb unseres vortheoretischen Überzeugungssystems einnimmt. Um sich die tatsächlichen Konsequenzen des Epiphänomenalismus aber vor Augen zu führen, ist es wichtig, einzelne Aspekte dieser Zentralität zu explizieren und möglicherweise zu hinterfragen. Was genau steht auf dem Spiel, wenn die Annahme mentaler Verursachung aufgegeben wird? Lassen sich die befürchteten Konsequenzen vermeiden?

Einige Autor*innen bestreiten, dass der Epiphänomenalismus um jeden Preis zu vermeiden ist. Sie behandeln den Epiphänomenalismus als ernstzunehmende theoretische Option, die zwar auf den ersten Blick unattraktiv wirken mag, sich aber bei genauerem Hinsehen als durchaus vertretbar erweist. In diesem Kapitel möchte ich die Aussichten dieser Option ausloten: Es geht also um die Frage, ob die Ablehnung der Annahme der Existenz mentaler Verursachung eine prinzipiell plausible Reaktion auf den Vorwurf des Epiphänomenalismus darstellt.

Ich gehe dabei wie folgt vor: In den ersten drei Abschnitten diskutiere ich Überlegungen, die für die Existenz mentaler Verursachung sprechen und gehe dabei auch auf Erwiderungen von Seiten des Epiphänomenalismus ein. Abschnitt 3.5.1. geht dabei auf ein Argument ein, demzufolge kausal ineffektive Ereignisse schlicht nicht existieren, so dass die Existenz mentaler Ereignisse von der Existenz mentaler Verursachung abhängt. In Abschnitt 3.5.2. geht es um ein Argument, das die Existenz von Handlungen an die Existenz mentaler Verursachung knüpft. In Abschnitt 3.5.3. führe ich anknüpfend an die vorigen Überlegungen eine Unterscheidung zwischen einem radikalen und einem konservativen Epiphänomenalismus ein. In Abschnitt 3.5.4. schildere ich eine allgemeine Strategie zur Verteidigung eines konservativen Epiphänomenalismus.

3.5.1 Mentale Verursachung und die Existenz mentaler Ereignisse

Der Epiphänomenalismus besagt, dass es mentale Ereignisse zwar gibt, sie aber keinerlei Wirkungen haben. Die Theorie steht in einem direkten Konflikt mit einem allgemeinem Prinzip, das diktiert, nur solche Entitäten zu postulieren,

is arguably a 'Moorean fact' – a fact so obvious that any philosophical view had better respect it if it hopes to be taken seriously".

die Wirkungen haben. Dieses Prinzip wird manchmal als ‚eleatisches Prinzip‘[123] oder als ‚Alexanders Diktum‘[124] bezeichnet. Wir können es für die vorliegenden Zwecke wie folgt formulieren:

Eleatisches Prinzip: Wenn ein Ereignis existiert, dann hat es Wirkungen.[125]

Das eleatische Prinzip besagt gewissermaßen, dass Wirksamkeit die Eintrittskarte zur Realität ist. Epiphänomenalist*innen müssen das eleatische Prinzip ablehnen. Denn im Epiphänomenalismus gibt es vollkommen wirkungslose Ereignisse. Zudem unterminiert dieses Prinzip die Unterscheidung zwischen Epiphänomenalismus und eliminativem Physikalismus: Die Wirkungslosigkeit mentaler Ereignisse kommt ihrer Elimination gleich. Es ist nicht möglich, zugleich die Wirkungslosigkeit und die Existenz mentaler Ereignisse zu verteidigen.

Zudem ergibt sich aus dem eleatischen Prinzip ein einfaches Argument für die Existenz mentaler Verursachung:

Das Argument aus dem eleatischen Prinzip

(1) Wenn es keine mentale Verursachung gibt, dann gibt es keine mentalen Ereignisse.
(2) Es gibt mentale Ereignisse.
(3) Also: Es gibt mentale Verursachung.

Offenbar müssen Epiphänomenalist*innen Prämisse (1) dieses Arguments mitsamt des eleatischen Prinzips, auf dem diese Prämisse beruht, zurückweisen. Wie steht es also um die Plausibilität des eleatischen Prinzips?

Das verbreitetste Argument für das eleatische Prinzip ist ein epistemisches Argument. Jens Harbecke etwa motiviert das eleatische Prinzip wie folgt:

---

[123] Armstrong (1978: 45–46) verweist auf Platons Sophistes. Der Gast aus Eleas sagt: „Ich setze nämlich als Erklärung fest, um das Seiende zu bestimmen, dass es nichts anderes ist als das was vermag.“ (Platon, Sophistes 274e).

[124] Kim (1993b: 202) verweist hierfür auf Samuel Alexanders (1927) Kritik des Epiphänomenalismus.

[125] Diese Formulierung beschränkt den Gehalt des Prinzips auf das für die vorliegende Argumentation Wesentliche. Erstens beschränkt es den Anwendungsbereich des Prinzips auf Ereignisse, während andere Formulierungen alle Entitäten betreffen. Zweitens ist es als Konditional und nicht wie andere Formulierungen als Bikonditional formuliert. Vgl. z. B. Kim (1993: 202) „To be real is to have causal powers“, Harbecke (2008: 81): „Everything that exists has causal effects“ und Colyvan (1998: 314): „An entity is to be counted as real if and only if it is capable of participating in causal processes“.

> The intuitive idea behind this principle is that there lies a severe problem in the introduction of entities into ontology that have nothing to do with what goes on causally in that ontology and that are not detectable in any way (if something can be detected, or perceived, then it at least causes certain perceptions). If such things as causally non-efficacious entities existed, there would be no way of knowing that they did.[126]

Die Idee ist hier also, dass unser epistemischer Zugang zu Entitäten davon abhängig ist, dass diese Entitäten einen kausalen Einfluss auf uns ausüben. Wir können zum Beispiel nur dann Kenntnis von der Existenz von Tischen erlangen, wenn Tische Wahrnehmungen von Tischen und Überzeugungen über Tische verursachen. Übertragen auf mentale Ereignisse bedeutet das: Wenn mentale Ereignisse keine Wirkungen haben, können wir nicht wissen, dass es sie gibt. Wir haben dann auch keinen Grund, ihre Existenz zu postulieren.

Es ist jedoch fraglich, ob sich diese Überlegung problemlos auf das Wissen über die eigenen mentalen Zustände übertragen lässt. Man kann ohne weiteres zugestehen, dass Wissen über äußere Gegenstände wie Tische impliziert, dass diese Gegenstände kausal wirksam sind. Im Epiphänomenalismus geht es jedoch um mentale Zustände. Und es scheint nicht plausibel, dass man nur dann Wissen von den eigenen mentalen Zuständen haben kann, wenn diese Wahrnehmungen oder Überzeugungen – also weitere mentale Zustände – verursachen. Dies würde ohnehin wenig erklären. Denn es würde sich nun die Frage stellen, wie wir Wissen von diesen weiteren mentalen Zuständen haben können. Diese müssten daher wiederum weitere mentale Zustände verursachen usw. Es scheint hier ein Regress zu drohen. Es ließe sich auf das epistemische Argument für das eleatische Prinzip also erwidern, dass der epistemische Zugang zu den eigenen mentalen Zuständen nicht – wie im Falle von Wahrnehmungswissen – kausal vermittelt ist. Wir können daher auch Wissen von den eigenen mentalen Zuständen haben, ohne dass sie kausal wirksam sind.[127]

Diese Erwiderung beruht auf einer epistemischen Sonderrolle mentaler Ereignisse. Sie kann daher im gegebenen Kontext nicht vollständig überzeugen. Denn die Exklusionsargumente, die den hier diskutierten Epiphänomenalismus motivieren, sind generalisierbar. Im Erfolgsfall zeigen sie nicht nur, dass mentale Ereignisse keine Wirkungen haben, sondern sie zeigen, dass höherstufige Ereignisse im Allgemeinen keine Wirkungen haben.[128] Die eben besprochene Erwiderung würde jedoch nur dazu berechtigen, die Existenz wirkungsloser mentaler Ereignisse zu postulieren. Insofern Epiphänomenalist*innen unter dem Einfluss der Exklusionsargumente

[126] Harbecke (2008: 81).

[127] Vgl. Robinson (2006: 89).

[128] Vgl. Block (2003).

aber einen allgemeinen Epiphänomenalismus über höherstufige Ereignisse vertreten wollen, müssten sie zum Beispiel auch die Existenz wirkungsloser biologischer oder neurologischer Ereignisse postulieren. Diese ist aber nach allem bis hierher Gesagten weiterhin ausgeschlossen. Wenn zum Beispiel biologische Ereignisse keine entsprechenden Wahrnehmungen oder Überzeugungen verursachen, haben wir keinen epistemischen Zugang zu ihnen und daher keinen Grund, ihre Existenz zu postulieren.

Eine allgemeinere Strategie zur Zurückweisung des epistemischen Arguments für das eleatische Prinzip besteht in einer prinzipiellen Zurückweisung der Annahme, dass wir nur Wissen von Ereignissen haben können, die in kausalem Kontakt mit uns stehen. Diese Annahme sollte nicht nur in Bezug auf mentale Ereignisse, sondern in Bezug auf zumindest alle höherstufigen Ereignisse zurückgewiesen werden.

Eine Frage, die Epiphänomenalist*innen hier aufwerfen können, ist, warum ausschließlich *kausaler* Kontakt einen epistemischen Zugang erlauben sollte. Selbst wenn biologische Ereignisse nicht selbst Überzeugungen verursachen, haben sie ja doch physische Basen, die Überzeugungen verursachen. Und dies führt dazu, dass wir zuverlässig Überzeugungen über biologische Ereignisse entwickeln, wenn biologische Ereignisse stattfinden. Warum sollte dies nicht für einen epistemischen Zugang zur Existenz biologischer Ereignisse ausreichen?

Diese (an dieser Stelle bloß angedeutete) Erwiderung ist eine Instanz einer allgemeinen Strategie zur Verteidigung des Epiphänomenalismus: Gegner*innen des Epiphänomenalismus behaupten, dass eine kausale Verbindung notwendig in einem bestimmten theoretischen Kontext ist: In diesem Fall soll eine kausale Verbindung notwendig sein, um einen epistemischen Zugang zur Existenz eines biologischen Ereignisses zu haben. Epiphänomenalist*innen können antworten, dass eine solche *kausale* Verbindung nicht notwendig ist. Stattdessen reicht eine *nicht-kausale* Verbindung, die sich aus der kausalen Verbindung zwischen der physischen Basis des in Frage stehenden Ereignisses und einem anderen Ereignis ergibt. In diesem Falle ließe sich etwa argumentieren, dass der Fakt, dass die Basis des biologischen Ereignisses entsprechende Überzeugungen verursacht, dazu führt, dass es eine stabile Verbindung zwischen dem Stattfinden biologischer Ereignisse und entsprechenden Überzeugungen gibt und dass dies für unseren epistemischen Zugang zu biologischen Ereignissen ausreicht.

3.5.2 Mentale Verursachung und die Existenz von Handlungen

Ein weiteres Argument für die Existenz mentaler Verursachung (und somit gegen den Epiphänomenalismus) beruft sich auf eine Verbindung zwischen der Existenz mentaler Verursachung und der Existenz von Handlungen. Wir können dieses Argument wie folgt rekonstruieren:

Das Argument aus der Existenz von Handlungen

(AH-1) Wenn es keine mentale Verursachung gibt, dann gibt es keine Handlungen.
(AH-2) Es gibt Handlungen.
(AH-K) Also: Es gibt mentale Verursachung.

Die Grundidee dieses Arguments ist, dass der Epiphänomenalismus falsch sein muss, weil er die absurde Konsequenz hat, dass niemand jemals eine Handlung ausführt. Diese Konsequenz würde unser gesamtes Menschenbild ins Wanken bringen. Denn wenn niemand jemals eine Handlung ausführt, ist auch niemand jemals für irgendetwas moralisch verantwortlich. Wir könnten niemals berechtigterweise eine Person für ihre Handlungen loben oder kritisieren. Schlimmer noch: Unsere gesamte Interpretation des Verhaltens anderer Menschen wäre falsch. Susi hebt nicht ihre Hand, um die Aufmerksamkeit der Dozentin auf sich zu lenken. Stattdessen hebt *sich* ihre Hand, und Susis vorhergehender Wunsch hat mit dieser Körperbewegung eigentlich nichts zu tun. Diese Überlegung macht durchaus plausibel, dass der Epiphänomenalismus, wie Fodor sagt, mit einem ‚Ende der Welt' einhergehen würde.[129]

Eine wichtige Motivation für Prämisse (AH-1) ist die kausale Handlungstheorie. Diese besagt grob gesprochen, dass Handlungen Ereignisse mit einer besonderen kausalen Geschichte sind: Handlungen sind genau jene Ereignisse, die von rationalisierenden mentalen Zuständen verursacht werden. Susis Handbewegung ist beispielsweise deshalb als Handlung zu zählen, weil sie von einem Wunsch und einer Überzeugung verursacht wurde, die es rational machen, die Hand zu heben.

[129] Vgl. z. B. Kim (1998: 31): „[T]he possibility of human agency evidently requires that our mental states – our beliefs, desires, and intentions – have causal effects in the physical world: in voluntary actions our beliefs and desires, or intentions and decisions, must somehow cause our limbs to move in appropriate ways, thereby causing the objects around us to be rearranged" und Ewing (1953: 128): „That epiphenomalism is false is assumed in all practical life. In deciding what we are to do we must always make the assumption that our will and motives can affect our physical actions, and it is silly to adopt a philosophy the denial of which is implied by us every time we do anything voluntarily".

Susi will die Aufmerksamkeit der Dozentin auf sich lenken und glaubt, dass das Armheben ein geeignetes Mittel dafür ist, dieses Ziel zu erreichen. Dieses rationalisierende Paar aus Überzeugung und Wunsch verursacht, dass Susi ihre Hand hebt. Hierdurch sind Handlungen von anderen Körperbewegungen unterschieden. Wenn Susi niest, gibt es kein passendes Paar aus Überzeugung und Wunsch, das die zugehörige Körperbewegung zugleich verursacht und sie rational erscheinen lässt. Deshalb handelt es sich auch nicht um eine Handlung.

Die Wahrheit von Prämisse (AH-1) des Arguments aus der Existenz von Handlungen folgt unmittelbar aus der kausalen Handlungstheorie. Wenn es keine mentale Verursachung gibt, dann gibt es keine Ereignisse, die durch rationalisierende mentale Zustände verursacht werden. Entsprechend gibt es dann auch keine Handlungen. Der Epiphänomenalismus zwingt daher, wie es scheint, zur Abkehr von der kausalen Handlungstheorie.

Auch hier können Epiphänomenalist*innen jedoch die Frage aufwerfen, warum nur eine *kausale* Verbindung zwischen mentalen Ereignissen und behavioralen Ereignissen eine zufriedenstellende Bestimmung des Handlungsbegriffs erlauben sollte. Warum sollten Handlungen nicht dadurch von bloßen Körperbewegungen unterschieden sein, dass sie von den physischen Basen rationalisierender mentaler Ereignisse verursacht werden? Denn wieder ergibt sich aus der kausalen Verbindung zwischen den Basen der relevanten mentalen Ereignisse und den behavioralen Ereignissen auch eine enge Verbindung zwischen den mentalen Ereignissen selbst und den behavioralen Ereignissen: So lässt sich plausibel argumentieren, dass behaviorale Ereignisse unter den Voraussetzungen des nicht-reduktiven Physikalismus von mentalen Ereignissen kontrafaktisch abhängen. Wenn dies aber so ist, lässt sich eventuell eine kontrafaktische Handlungstheorie als Ersatz für eine kausale Handlungstheorie verteidigen. Wieder können Epiphänomenalist*innen also argumentieren, dass die theoretische Arbeit, die laut Kausalist*innen nur von der Verursachungsrelation geleistet werden kann, im Rahmen eines Epiphänomenalismus von einer nicht-kausalen Ersatzrelation übernommen werden kann. Der Epiphänomenalismus hätte dann keinen Nachteil gegenüber dem Kausalismus.

3.5.3 Radikaler und konservativer Epiphänomenalismus

Eine Reihe von Einwänden gegen den Epiphänomenalismus haben eine gemeinsame Struktur: Es wird zunächst argumentiert, dass der Epiphänomenalismus eine bestimmte, scheinbar absurde Konsequenz hat. Es wird dann behauptet, dass diese

absurde Konsequenz abgelehnt werden sollte. Daraus folgt, dass der Epiphänomenalismus falsch ist. Diese allgemeine Struktur kann wie folgt festgehalten werden:

Argument gegen den Epiphänomenalismus

(1) Wenn es keine mentale Verursachung gibt, dann folgt die absurde Konsequenz a.

(2) Die absurde Konsequenz a ist falsch.

(3) Also: Es gibt mentale Verursachung.

Für ‚a' kann nun etwa eingesetzt werden: ‚Es gibt keine mentale Ereignisse', ‚Wir haben kein Wissen von unseren eigenen mentalen Zuständen', ‚Es gibt keine Handlungen', ‚Es gibt keine mentale Verantwortlichkeit' usw.

Diese Rekonstruktion macht noch einmal deutlich, dass es prinzipiell zwei verschiedene Weisen gibt, den Epiphänomenalismus gegen solcherart Einwände zu schützen: Auf der einen Seite gibt es die konservative Reaktion: Es wird bestritten, dass der Epiphänomenalismus eine solche Konsequenz tatsächlich hat. Auf der anderen Seite gibt es die radikale Reaktion: Es wird die (scheinbar) absurde Konsequenz in Kauf genommen. Im Falle des Arguments aus der Existenz von Handlungen besteht die konservative Reaktion darin, die Verbindung zwischen Handlungen und mentaler Verursachung zu bestreiten. Die radikale Reaktion auf das Argument aus der Existenz von Handlungen würde hingegen darin bestehen, die Existenz von Handlungen ganz einfach abzustreiten und hieraus die Konsequenz zu ziehen, dass unsere alltäglichen Erklärungen des Verhaltens tatsächlich durchgehend falsch sind und tatsächlich niemand jemals für etwas moralisch verantwortlich ist. Diese Option wird selten in all ihrer Radikalität gewählt.[130]

Man kann entsprechend eine (grobe und graduelle) Unterscheidung zwischen konservativen und radikalen Versionen des Epiphänomenalismus ziehen. Die allermeisten Epiphänomenalist*innen sind tendenziell konservative Epiphänomenalist*innen: Sie meinen, dass der Epiphänomenalismus keine radikalen

[130] Hommen (2013: 281–308) argumentiert auf der Grundlage des Epiphänomenalismus für die Falschheit kausaler Handlungserklärungen, meint aber, dass daraus zumindest nicht folgt, dass es keine handelnden Personen gibt. Auch er kommt daher zu dem Schluss, dass der Epiphänomenalismus keine radikalen Konsequenzen für unser Menschenbild hat (vgl. z. B. deutlich Hommen (2013: 312)).

Konsequenzen für unser vortheoretisches Menschen- und Weltbild hat.[131] Im Prinzip wäre aber auch ein radikaler Epiphänomenalismus denkbar: Eine Position, die den Epiphänomenalismus mitsamt all der radikalen Konsequenzen für unser Menschen- und Weltbild in voller Schärfe akzeptiert.

Zugleich wird klar, dass Fodors Bedenken vom ‚Ende der Welt' den Epiphänomenalismus letztlich nur dann treffen kann, wenn ein *konservativer* Epiphänomenalismus nicht möglich ist. Denn ein konservativer Epiphänomenalismus ist gerade dadurch ausgezeichnet, dass er das ‚Ende der Welt' vermeidet. Solange also ein konservativer Epiphänomenalismus doch verteidigt werden kann, ist eine ‚Epiphobie' eigentlich nicht angebracht. Der Epiphänomenalismus ist dann nichts, wovor man sich fürchten müsste.

Um Fodors Metapher der Epiphobie weiterzuführen, kann man diesen Punkt auch wie folgt fassen: Ängste können auf mindestens zweierlei Weisen irrational sein. Einerseits kann eine Angst irrational sein, weil der Gegenstand der Angst nicht einzutreten droht. Ein Beispiel wäre die Angst der Figur Majestix aus den Asterix-Comics davor, dass ihm der Himmel auf den Kopf fällt. Zwar wäre es schlimm, wenn einem der Himmel auf den Kopf fallen würde. Das aber ist nichts, was einzutreten droht und daher nichts, wovor man sich fürchten müsste. Andererseits kann eine Angst irrational sein, weil der Gegenstand der Angst harmlos ist. Ein Beispiel wäre hier die Angst einer Person, die sich für einen Vampir hält und daher Angst vor dem Sonnenaufgang hat. Der Sonnenaufgang wird unvermeidlich eintreten, aber zieht keinerlei Schäden mit sich. Er ist daher ebenfalls nichts, wovor man sich fürchten müsste.

Kausalist*innen meinen typischerweise, dass die Epiphobie mit Majestix' Angst davor, dass ihm der Himmel auf den Kopf fällt, vergleichbar ist. Der Epiphänomenalismus ist der Sache nach fürchtenswert. Jedoch gibt es – auch aus Sicht des nicht-reduktiven Physikalismus – keine guten Gründe, an ihn zu glauben. Konservative Epiphänomenalist*innen würden die Epiphobie eher mit der Angst vor dem Sonnenaufgang vergleichen. Der Epiphänomenalismus folgt tatsächlich aus dem nicht-reduktiven Physikalismus. Allerdings folgen aus dem Epiphänomenalismus keine schrecklichen Konsequenzen. Radikale Epiphänomenalist*innen schließlich würden die Epiphobie gerade nicht für irrational halten. Denn der Epiphänomenalismus ist zugleich wahr und wahrhaft schrecklich. Er ist dann eine unangenehme Wahrheit, mit der zu leben wir lernen müssen.

---

[131] Eliminative Physikalist*innen oder manche Kritiker*innen der Existenz von Willensfreiheit sind hier deutlich radikaler eingestellt.

3.5.4 Die Strategie der Ersatzrelation und die Metaphysik der Verursachung

Verteidigungen des konservativen Epiphänomenalismus folgen einem gemeinsamen Muster: Auch, wenn mentale Ereignisse nicht im Wortsinne behaviorale Ereignisse verursachen, gibt es dennoch enge Verbindungen zwischen mentalen Ereignissen und behavioralen Ereignissen. Denn behaviorale Ereignisse werden zwar nicht durch die mentalen Ereignisse selbst verursacht. Sie werden jedoch durch die physischen Basen der mentalen Ereignisse verursacht. Und das wiederum führt dazu, dass es nomologische oder kontrafaktische Verbindungen zwischen mentalen und behavioralen Ereignissen gibt. Diese nomologischen oder kontrafaktischen Verbindungen können nun quasi für die Kausalrelation einspringen und alle Funktionen übernehmen, die laut den Gegner*innen des Epiphänomenalismus durch die Kausalrelation übernommen wird.

Die Verteidigung des konservativen Epiphänomenalismus beruht also auf der Annahme, dass es eine geeignete *Ersatzrelation* gibt, die die vermeintliche theoretische Arbeit der Kausalrelation übernehmen kann. Gegeben, dass sich eine solche Ersatzrelation finden lässt, kann die Annahme, dass es mentale *Verursachung* gibt, guten Gewissens aufgegeben werden.

In Bezug auf unser Beispiel heißt dies folgendes: Zwar verursacht Susis Wunsch nicht selbst, dass Susi ihren Arm hebt. Jedoch ist Susis Wunsch ontologisch abhängig von einem physischen Ereignis, das verursacht, dass sie ihren Arm hebt. Dies führt dazu, dass ähnliche Armbewegung regelmäßig auf ähnliche Wünsche folgen, und dass Susi nicht ihren Arm gehoben hätte, wenn sie nicht einen entsprechenden Wunsch gehabt hätte. Insofern solche regelmäßigen Abfolgen oder kontrafaktischen Verbindungen keine kausalen Verbindungen implizieren, ist all dies mit dem Epiphänomenalismus kompatibel. Und insofern solche nomologischen oder kontrafaktischen Verbindungen eine Vermeidung der vermeintlichen absurden Konsequenzen des Epiphänomenalismus erlauben, kann der Epiphänomenalismus gegen die erwähnten Einwände verteidigt werden.

Ein Kandidat für eine solche Ersatzrelation ist wie gesagt – geeignet qualifizierte – kontrafaktische Abhängigkeit. Man betrachte etwa die folgenden beiden Zitate:

> [T]he objections against epiphenomenalism seem to be based on the widely shared assumption that physical events like behaviour do not only follow regularly upon mental events, but that their occurrence depends on the occurrence of the latter ('he wouldn't have cried if he hadn't felt pain'). Here the proponent of epiphenomenalism can call attention to the fact that she can happily admit these kinds of dependencies,

> which are, furthermore, also part of her explanation of the possibility of knowledge of mental states.[132]

> It is generally recognized that counterfactual dependence is weaker than causal dependence. What is not generally recognized is that this fact undermines intuitive arguments against epiphenomenalism, for the relevant intuitions may be about counterfactual, rather than causal, dependence.[133]

Die Strategie, sich auf kontrafaktische Abhängigkeit als Ersatzrelation zu berufen, hat zwei wichtige Voraussetzungen:

Erstens muss sich (geeignet qualifizierte) kontrafaktische Abhängigkeit dafür eignen, den Epiphänomenalismus von seinen vermeintlichen, absurden Konsequenzen zu trennen. Im Kontext des Arguments aus der Existenz von Handlungen heißt das mindestens, dass die kontrafaktische Abhängigkeit behavioraler Ereignisse von passenden mentalen Ereignissen eine geeignete Grundlage für die Unterscheidung zwischen Handlungen und bloßen Ereignissen darstellt. Die Funktion, die in der kausalen Handlungstheorie von der Kausalrelation übernommen wird, kann dann von der Relation der (geeignet qualifizierten) kontrafaktischen Abhängigkeit als Ersatzrelation übernommen werden.

Zweitens muss kontrafaktische Abhängigkeit aber auch von der Kausalrelation getrennt werden. Denn anderenfalls wäre die resultierende Position eben keine Version des Epiphänomenalismus, sondern eine Version des Kausalismus. Epiphänomenalist*innen werden daher gewissermaßen dazu gedrängt, eine Theorie der Kausalität zu vertreten, der zufolge Verursachung nicht bloß in – geeignet qualifizierter – kontrafaktischer Abhängigkeit besteht. Allgemein gilt: Je mehr von einer ‚genuinen' Kausalrelation verlangt ist, desto mehr Raum bleibt Epiphänomenalist*innen, sich auf eine nicht-kausale Ersatzrelation zu beziehen. Wenn beispielsweise kontrafaktische Abhängigkeit zwischen gänzlich distinkten zeitlich aufeinanderfolgenden Ereignissen – anders als typische Vertreter*innen kontrafaktischer Theorien behaupten – *nicht* hinreichend für Kausalität ist, dann ist eben diese Relation eine mögliche nicht-kausale Ersatzrelation, auf die sich (konservative) Epiphänomenalist*innen zur Verteidigung ihrer Position berufen können. In einem gewissen Sinne machen anspruchsvolle Kausalbegriffe die Aufgabe für konservative Epiphänomenalist*innen daher leichter.

Es bleibt also, die folgenden Punkte festzuhalten:

---

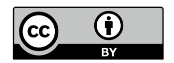
[132] Pauen, Staudacher und Walter (2006: 13).

[133] Lyons (2006: 779). Lyons Argument für den Epiphänomenalismus ist deutlich subtiler und kann hier nicht dargestellt werden.

Erstens: Der Epiphänomenalismus und der Kausalismus sind oft nicht so weit voneinander entfernt, wie man denken könnte. Wenn etwa Epiphänomenalist*innen auch geeignet qualifizierte Beziehungen kontrafaktischer Abhängigkeit zwischen mentalen Ereignissen und ihren (bloß) vermeintlichen Wirkungen akzeptieren dürfen und auf dieser Grundlage Argumente gegen den Epiphänomenalismus zurückweisen können, dann unterscheidet sich ihre Position kaum noch von der Position von manchen erklärten Kausalist*innen, die sich ebenfalls auf (geeignet qualifizierte) Beziehungen kontrafaktischer Abhängigkeit berufen und hieraus aber Kausalbehauptungen folgern.[134] Die Entscheidung, ob man eine kausalistische oder eine epiphänomenalistische Erwiderung auf den Vorwurf des Epiphänomenalismus vertritt, wird in diesem Fall unter Umständen zu einer bloß terminologischen Entscheidung: Wer meint, dass geeignet qualifizierte kontrafaktische Abhängigkeit Kausalität impliziert, und die Existenz mentaler Verursachung auf dieser Grundlage verteidigt, ist Kausalist*in. Wer hingegen meint, dass Kausalität nicht aus (qualifizierter) kontrafaktischer Abhängigkeit folgt, aber den konservativen Epiphänomenalismus auf Basis kontrafaktischer Abhängigkeiten verteidigt, ist Epiphänomenalist*in. Beide Positionen sind aber in ihren substantiellen Annahmen identisch: In beiden Fällen bestehen bestimmte Beziehungen kontrafaktischer Abhängigkeit zwischen behavioralen und mentalen Ereignissen. Und in beiden Fällen sind es diese kontrafaktischen Abhängigkeiten, die absurde Konsequenzen des nicht-reduktiven Physikalismus verhindern.

Zweitens: Die Nähe mancher Versionen des Kausalismus zu manchen Versionen des Epiphänomenalismus bedeutet nicht unbedingt, dass diese Versionen des Kausalismus um jeden Preis zu vermeiden sind. Worauf es eigentlich ankommt, ist, bestimmte absurde Konsequenzen des nicht-reduktiven Physikalismus zu vermeiden. Insofern dies sowohl konservativen Versionen des Epiphänomenalismus als auch dem Kausalismus gelingt, sind die beiden Erwiderungen auf den Vorwurf des Epiphänomenalismus auf einer Stufe. Es gibt unter dieser Voraussetzung keinen prinzipiellen Grund, den Kausalismus dem Epiphänomenalismus vorzuziehen.

[134] Vgl. Loewer (2002), (2007) und (2015).


cc
BY

4 Mentale Verursachung und Kausalität

Bis hierher habe ich sowohl den nicht-reduktiven Physikalismus als auch den Vorwurf des Epiphänomenalismus weitgehend unabhängig von der Frage behandelt, was Verursachung ist. Wie in Teil 3 dieser Arbeit jedoch an einigen Stellen deutlich wurde, ist diese Frage von großer Relevanz für die Plausibilität des Vorwurfs des Epiphänomenalismus: Die These der kausalen Geschlossenheit, das übliche Exklusionsprinzip sowie die drei Prinzipien der Verursachung im Stufenmodell der Realität sind in ihrer Plausibilität ein Stück weit davon abhängig, welcher Kausalbegriff ihnen zugrunde gelegt wird. Zudem sind auch die Aussichten einer epiphänomenalistischen Erwiderung auf den Vorwurf des Epiphänomenalismus von der Frage abhängig, was Kausalität ist.

Es ist daher naheliegend, sich etwas detaillierter mit der Philosophie der Kausalität auseinanderzusetzen, um sich der Frage nach der Schlagkräftigkeit des Vorwurfs des Epiphänomenalismus anzunähern. In diesem dritten Teil der Arbeit geht es daher um den Zusammenhang zwischen verschiedenen Fragestellungen aus der Philosophie der Kausalität auf der einen Seite und dem Vorwurf des Epiphänomenalismus auf der anderen Seite.

Mein Vorgehen wird folgendes sein: In Abschnitt 4.1. führe ich ein Verständnis von Kausalität als kausaler Abhängigkeit ein und grenze es von einem Verständnis von Kausalität als kausaler Produktion und dem in Teil 3 bereits mehrfach zur Sprache gekommenen Begriff der hinreichenden Verursachung ab. Abschnitt 4.2. geht knapp auf die Rolle hinreichender Verursachung im Vorwurf des Epiphänomenalismus ein.

In den Abschnitten 4.3. und 4.4. – die zusammen gewissermaßen den Kern des Teils 4 dieser Arbeit ausmachen – setze ich mich detailliert mit zwei verbreiteten Theorien kausaler Abhängigkeit und ihrer Anwendung auf mentale Verursachung im nicht-reduktiven Physikalismus auseinander: Mit der kontrafaktischen Theorie

M. Rolffs, *Kausalität und mentale Verursachung*, Abhandlungen zur Philosophie,
https://doi.org/10.1007/978-3-662-66778-1_4

der Kausalität und der interventionistischen Theorie der Kausalität. Ich argumentiere dafür, dass beide Theorien es erlauben, den ersten Schritt des Vorwurfs des Epiphänomenalismus zurückzuweisen. Das Ergebnis dieser beiden Abschnitte ist also, dass der nicht-reduktive Physikalismus nicht zur Konsequenz hat, dass es keine mentale kausale Abhängigkeit gibt.

Nehmen wir die in 4.1. eingeführte Unterscheidung zwischen kausaler Abhängigkeit und kausaler Produktion jedoch ernst, ist der Vorwurf des Epiphänomenalismus mit dieser Einsicht noch nicht vollständig zurückgewiesen. Denn insofern es kausale Produktion als weitere Kausalrelation gibt, stellt sich die Frage, ob aus dem nicht-reduktiven Physikalismus folgt, dass es keine kausale Produktion gibt. Auf diese Problematik gehe ich im letzten Abschnitt 4.5. ein. Hier führe ich einige Überlegungen an, die dafür sprechen, dass nicht-reduktive Physikalist*innen einen Epiphänomenalismus in Bezug auf *mentale kausale Produktion* akzeptieren können, insofern sie einen Epiphänomenalismus in Bezug auf *mentale kausale Abhängigkeit* vermeiden können.

4.1 Kausalität als kausale Abhängigkeit

Dieser Abschnitt dient einer vorläufigen Charakterisierung und Abgrenzung der Relation kausaler Abhängigkeit. Ich gehe wie folgt vor:

In Abschnitt 4.1.1. wird die Unterscheidung zwischen kausaler Abhängigkeit und kausaler Produktion erläutert. Abschnitt 4.1.2. betrifft das Verhältnis zwischen kausaler Abhängigkeit und hinreichender Verursachung. In Abschnitt 4.1.3. gehe ich auf die Idee der proportionalen Verursachung und ihrer Beziehung zu kausaler Abhängigkeit ein.

4.1.1 Kausale Abhängigkeit und kausale Produktion

In seinem auch in der Debatte um den Vorwurf des Epiphänomenalismus stark rezipierten[1] Aufsatz ‘Two Concepts of Causation’ trifft Ned Hall die folgende Unterscheidung zwischen zwei Begriffen der Kausalität:

> Causation, understood as a relation between events, comes in at least two basic and fundamentally different varieties. One of these, which I call “dependence”, is simply that: counterfactual dependence between wholly distinct events. […] The second variety is rather more difficult to characterize, but we evoke it when we say of an event

[1] Vgl. z. B. die Debatte zwischen Loewer (2007) und Kim (2007).

> c that it helps to generate or bring about or produce another event e, and for that reason I call it "production."[2]

Demzufolge gibt es also einerseits einen Kausalitätsbegriff – *kausale Abhängigkeit*[3] – der in einem engen Zusammenhang mit der Idee steht, dass Ursachen einen Unterschied für ihre Wirkungen machen und der laut Hall angemessen über kontrafaktische Abhängigkeit spezifiziert werden kann. Andererseits gibt es einen anderen Kausalitätsbegriff – *kausale Produktion* – der in einem engen Zusammenhang zu der Idee steht, dass Ursachen ihre Wirkungen hervorbringen und – um Jaegwon Kims schon in Abschnitt 3.4.3. erwähnte Metapher zu bemühen – 'kausale Arbeit' leisten. Es handelt sich hierbei laut Hall um distinkte Begriffe, die in einigen Fällen auseinandergehen können.

Was macht kausale Produktion genauer aus? Ein wichtiger Aspekt der Produktionsrelation wird durch eine *Lokalitätsbedingung* eingefangen. Eine solche findet sich zum Beispiel bei David Hume:[4]

> I find, in the first place, that whatever objects are consider'd as causes or effects, are *contiguous*. Tho' distant objects may sometimes seem productive of each other, they are commonly found upon examination to be link'd by a chain of causes, which are contiguous among themselves, and to the distant objects; and when in any particular instance we cannot discover this connexion, we still presume it to exist. We may therefore consider the relation of CONTIGUITY as essential to that of causation.[5]

Die Idee ist hier, dass Ursachen mit ihren Wirkungen über raumzeitlich kontinuierliche Kausalketten verbunden sind. Wenn ich beim Billard die weiße Kugel anstoße und einige Sekunden später die volle 11er Kugel in die rechte Mitteltasche fällt, ist mein Stoß eine Ursache dafür, dass die Kugel in die Tasche fällt. Diese Ursache ist über raumzeitlich benachbarte kausale Zwischenglieder mit der Wirkung verbunden: Mein Stoß zu t_0 verursacht zunächst, dass die weiße Kugel zu t_1 etwas weiter entfernt von ihrer Ausgangsposition ist und sich mit einer bestimmten Geschwindigkeit in Richtung der 11er Kugel bewegt. Dies verursacht, dass die Kugel zu t_2 mit der vollen 11er Kugel zusammenstößt, was wiederum verursacht, dass die 11er Kugel sich zu t_3 von ihrer Ausgangsposition

[2] Hall (2004: 225).

[3] Ich verwende 'kausale Abhängigkeit' hier anders als Lewis (1973b).

[4] Hiermit soll Hume natürlich nicht als ein klassischer Vertreter einer Produktionsauffassung von Kausalität eingeführt werden. Insofern Hume eine Regularitätstheorie der Kausalität vertreten hat, ist seine Theorie gerade keine Produktionsauffassung. Dieser Aspekt der Lokalitätsbedingung geht jedoch in die Richtung einer Produktionsauffassung.

[5] Hume (1739: 75) Vgl. auch Beebee (2006: 132–136) für Diskussion.

entfernt hat und zu t_4 schließlich in die Tasche fällt. Wir haben hier also eine (in dem Beispiel nur schemenhaft ausgeführte) Kette von raumzeitlich benachbarten Ursachen.[6]

Eng mit dieser Lokalitätsbedingung zusammen hängt die Idee, dass Ursachen mit ihren Wirkungen über physische Prozesse verknüpft sind. Eine theoretische Entwicklung dieser Idee findet sich etwa in Phil Dowes Erhaltungsgrößentheorie der Kausalität.[7] Laut Dowe sind Ursachen mit ihren Wirkungen über *kausale Prozesse* und *kausale Interaktionen* verknüpft – wobei kausale Prozesse als Weltlinien von Objekten verstanden werden, die eine Erhaltungsgröße haben und kausale Interaktionen als Überschneidungen der Weltlinien von Objekten verstanden werden, die mit einem Austausch einer Erhaltungsgröße verbunden sind.[8] Auch diese Idee ist in ihrer Anwendung auf das Billard-Beispiel sehr plausibel: Ich verleihe der weißen Billardkugel durch meinen Stoß einen bestimmten Impuls – eine Erhaltungsgröße. Die Weltlinie der weißen Kugel ist ein kausaler Prozess. Die Weltlinie der weißen Kugel überschneidet sich mit der Weltlinie der 11er Kugel. Hier wird der Impuls (oder ein Teil des Impulses) der weißen Kugel auf die 11er Kugel unter Impulserhaltung übertragen. Die mit Impuls versehene 11er Kugel bewegt sich, bis sie in die Tasche fällt. Auch die Weltlinie der 11er Kugel ist ein kausaler Prozess. Die Ursache ist hier also über kausale Prozesse und eine kausale Interaktion mit der Wirkung verbunden.

Kausale Produktion kann in einer ersten Annäherung über die erwähnten Redewendungen (Hervorbringen, kausale Arbeit usw.), die Lokalitätsbedingung und die Idee einer physischen Verknüpfung erläutert werden.[9] All dies sind sicherlich Aspekte, die für unser Alltagsdenken über Kausalität eine wichtige Rolle spielen. Jedoch gibt es Beispiele, in denen wir dazu geneigt sind, einen Faktor als Ursache für einen anderen Faktor zu klassifizieren, ohne dass

---

[6] Hall betrachtet die Lokalitätsbedingung nicht als im begrifflichen Sinne notwendige Bedingung für Kausalität, sondern als interessante kontingente Wahrheit (Vgl. Hall (2002: 276)). Das Problem mit den unten besprochenen Fällen von negativer Kausalität und Doppelverhinderung ist dann, dass sie eine Verletzung der Lokalitätsbedingung zu einfach etablieren: Sie würden zu „non-locality on the cheap“ führen (Hall (2002)).

[7] Der Verweis auf Dowes Theorie bietet sich im vorliegenden Kontext auch deshalb an, weil sich Kim (2007: 236) im Zusammenhang mit seiner Auffassung von kausaler Produktion auf diese Theorie bezieht. Andere Theorien der Kausalität, die von der Idee von Kausalität als Produktion ausgehen, finden sich zum Beispiel bei Salmon (1984) und Mumford und Anjum (2009).

[8] Vgl. Dowe (2000).

[9] Hall (2004) verweist weiterhin auf eine Intrinsitätsbedingung und auf eine Transitivitätsbedingung, die beide auf Kausalität als Produktion eingeschränkt werden.

sich eine lokale physische Verknüpfung zwischen Ursache und Wirkung findet. Wichtig sind hier insbesondere Fälle von negativer Kausalität und Fälle von Doppelverhinderung.

Hier ist ein typisches Beispiel für negative Kausalität:[10]

> **Negative Kausalität**: Mein Nachbar Peter hat versprochen, sich um meine Pflanzen zu kümmern, während ich im Urlaub bin. Das hat er aber nicht getan. Deshalb sind meine Pflanzen eingegangen.

Peters Unterlassung, meine Pflanzen zu gießen, ist plausiblerweise eine Ursache dafür, dass meine Pflanzen gestorben sind. Jedoch ist es wenig plausibel, dass Peters Unterlassung das Eingehen der Pflanzen *produziert* hat. Denn es gibt keine lokale physische Verknüpfung, die die Unterlassung mit dem Eingehen der Pflanzen verbindet: Eine raumzeitlich kontinuierliche Kette von kausalen Zwischengliedern ist hier nicht ohne weiteres zu finden.[11] Peter hat sich in der Zeit, in der er meine Pflanzen hätte gießen können, an allen möglichen Orten aufgehalten. In der Nähe meiner Pflanzen war er jedoch nicht. Und wenn wir in der raumzeitlichen Umgebung meiner Pflanzen nach Ursachen suchen, werden wir vielleicht feststellen, dass der Wasserdruck im Pflanzengewebe abgenommen hat und sie nach und nach vertrocknet sind. Peters Unterlassung werden wir dort aber nicht entdecken. Die Lokalitätsbedingung scheint auf diesen Fall keine plausible Anwendung zu haben. Auch gibt es keinen physischen Prozess, der Peters Unterlassung mit dem Tod der Pflanzen verbindet.

Peters Unterlassung scheint hier also nicht deshalb als Ursache für den Tod der Pflanzen zu zählen, weil sie den Tod der Pflanzen über einen lokalen physischen Prozess hervorgebracht hat. Denn dies ist gar nicht der Fall. Vielmehr zählt die Unterlassung einfach deshalb als Ursache, weil sie *einen Unterschied* für den

---

[10] In diesem Beispiel tritt ein negatives Ereignis als Ursache in Erscheinung. ‚Negative Kausalität' wird häufig so verstanden, dass auch Fälle von Verhinderung – in denen ein negatives Ereignis als Wirkung in Erscheinung tritt – als Fälle von negativer Kausalität zählen. Der Tod der Pflanzen selbst könnte – insofern er als eine Abwesenheit von Stoffwechselprozessen o.ä. bestimmt wird – ebenfalls als negatives Ereignis aufgefasst werden.

Bei einem noch weiteren Verständnis des Ausdrucks ‚negative Kausalität' zählen auch Fälle von Kausalität, bei denen weder Ursache noch Wirkung negative Ereignisse sind, als Fälle von negativer Kausalität, insofern Ursache und Wirkung durch kausale Zwischenglieder verbunden sind, die negative Ereignisse sind. Unter diesem Verständnis könnte auch der gleich diskutierte Fall von Doppelverhinderung als ein Fall von negativer Kausalität zählen.

[11] Gegen die Möglichkeit, eine geeignete Kette von raumzeitlich benachbarten negativen Ereignissen zu konstruieren, um an einer Lokalitätsbedingung festzuhalten, argumentiert Hall (2002: 278–288).

Tod der Pflanzen gemacht hat: Hätte Peter die Pflanzen gegossen, wären sie nicht eingegangen.[12]

Eine verwandte Klasse von Beispielen für intuitive Kausalbeziehungen, in denen wir keine Verknüpfung über einen lokalen physischen Prozess finden, besteht in Fällen von Doppelverhinderung:

> **Doppelverhinderung**: Susi klaut ein Stoppschild an einer vielbefahrenen Kreuzung. Später am Tag gibt es deshalb einen Unfall an der Kreuzung. Zwei Autos stoßen zusammen. Wenn Susi das Stoppschild nicht geklaut hätte, hätte das Stoppschild dafür gesorgt, dass eines der Autos stehenbleibt. Es wäre dann nicht zum Unfall gekommen.[13]

In diesem Beispiel verhindert Susis Diebstahl, dass die Anwesenheit des Stoppschilds den Unfall verhindert. In diesem Sinne haben wir es hier mit einem Fall von Doppelverhinderung zu tun.

In diesem Fall ist es äußerst plausibel, Susis Diebstahl als eine Ursache für den Unfall zu klassifizieren. Schließlich wäre es ohne den Diebstahl nicht zum Unfall gekommen. Susis Diebstahl hat also einen Unterschied dafür gemacht, ob der Unfall stattfindet. Zugleich ergeben sich hier jedoch ähnliche Probleme mit Blick auf die lokale physische Verknüpfung zwischen Diebstahl und Unfall, wie im obigen Fall von negativer Kausalität: Eine Kette von raumzeitlich benachbarten Ursachen, die den Diebstahl mit dem Unfall verbindet, ist nicht ohne weiteres zu finden.[14] Und schauen wir uns die lokalen physischen Prozesse an, die zum

---

[12] Ein Problem im Zusammenhang mit negativer Kausalität besteht darin, dass neben Peter auch weitere Personen meine Pflanzen nicht gegossen haben und der Tod der Pflanzen auch von ihren Unterlassungen kontrafaktisch abhängt. Intuitiv wollen wir aber zum Beispiel Barack Obamas Unterlassung, meine Pflanzen zu gießen, nicht als Ursache für den Tod der Pflanzen zählen. Es müsste also ein Kriterium dafür geben, negative Ereignisse, die als Ursachen zählen, von negativen Ereignissen zu unterscheiden, die nicht als Ursachen zählen. Und ein einfaches kontrafaktisches Kriterium ist nicht geeignet, diese Unterscheidung zu ziehen: Der Tod meiner Pflanzen hängt sowohl von der Unterlassung meines Nachbarn als auch von der Unterlassung Barack Obamas kontrafaktisch ab (Vgl. Beebee (2004), McGrath (2005) und Birnbacher und Hommen (2012) für weitere Diskussion). Dieses allgemeine Problem – das ich hier nicht lösen kann – sollte aber nicht darüber hinwegtäuschen, dass zumindest manche negativen Ereignisse recht eindeutig als Ursachen zu klassifizieren sind. Und hier haben wir dann Fälle von Kausalbeziehungen, in denen Ursache und Wirkung nicht durch lokale Prozesse miteinander verbunden sind.

[13] Das Beispiel übernehme ich von Hitchcock (2001: 362–362). Hall (2004: 241–242) verwendet ein strukturgleiches Beispiel.

[14] Hall (2002: 278–288) kritisiert die Möglichkeit, eine lokale kausale Kette aus Auslassungen zu konstruieren.

Unfall geführt haben, ist es naheliegend, zum Beispiel die Weltlinien der beiden zusammengestoßenen Fahrzeuge zu betrachten. Dabei ist aber von Susis Diebstahl keine Spur. Wieder scheint Susis Diebstahl also nicht deshalb als Ursache für den Unfall zu zählen, weil er über einen lokalen kausalen Prozess mit dem Unfall verbunden ist. Vielmehr ist allein die Beobachtung, dass Susis Diebstahl einen Unterschied macht, für das entsprechende Kausalurteil verantwortlich.

Wenn kausale Produktion also über die Lokalitätsbedingung und über die Idee der physischen Verknüpfung verstanden werden kann, liegt im Fall von negativer Kausalität keine Produktionsbeziehung zwischen der Unterlassung meines Nachbarn und dem Tod der Pflanzen vor. Im Fall von Doppelverhinderung liegt ebenfalls keine Produktionsbeziehung zwischen dem Diebstahl des Stoppschilds und dem Unfall vor. Insofern wir in diesen Fällen aber dennoch von Kausalbeziehungen sprechen möchten, ist der hierbei angewendete Begriff von Kausalität nicht begrifflich an Produktion gebunden. Wir haben es vielmehr mit einer Art von Kausalität zu tun, die ohne eine lokale physische Verknüpfung auskommt.

In Anbetracht dieser Situation stehen nun mehrere Optionen offen:

Erstens könnte man schließen, dass kausale Produktion entgegen dem ersten Anschein eben nicht wesentlich für Kausalität ist. Denn Kausalität kommt auch ohne kausale Arbeit und lokale physische Verknüpfung vor.[15]

Zweitens könnte man an der Idee, dass eine Lokalitätsbedingung, eine physische Verknüpfung und das Leisten kausaler Arbeit wesentlich für Kausalität sind, festhalten. In diesem Falle müsste man schließen, dass es keine Fälle von negativer Kausalität gibt: Die Unterlassung meines Nachbarn hat nicht verursacht, dass die Pflanzen eingegangen sind. Auch Fälle von Doppelverhinderung werden dann nicht als Fälle von genuiner Kausalität klassifiziert: Susis Diebstahl hat den Unfall nicht verursacht.[16]

Beide Optionen werden von unterschiedlichen Autor*innen in der Philosophie der Kausalität verfolgt. Tatsächlich weisen die geschilderten Fälle von negativer Kausalität und Doppelverhinderung auf eine Spaltung zwischen verschiedenen

---

[15] Vgl. z. B. Schaffer (2004b: 197): “Negative causation is genuine causation […]. It involves no physical connection between cause and effect. Thus causes need not be physically connected to their effects.”

[16] Vgl z. B. Mumford und Anjum (2009: 280): “Lewis accepts these consequences: that causation need not be an intrinsic matter, that there need be no continuous chain from cause to effect, that there can be causation by absence, and that there can be causation at a distance. It shall be argued that none of these are features we should want in genuine causation and that, if they follow from granting that double prevention is causation, then we should not do so.”

Ansätzen in der Philosophie der Kausalität hin: Auf der einen Seite gibt es Theorien der Kausalität, die der Lokalitätsbedingung und der Idee einer physischen Verknüpfung zwischen Ursache und Wirkung eine wichtige Rolle zugestehen und deren Verteidiger*innen gegen die Möglichkeit von negativer Kausalität oder Kausalität durch Doppelverhinderung argumentieren. Solche Ansätze können *Produktionstheorien der Kausalität* genannt werden. Auf der anderen Seite gibt es Theorien der Kausalität, die ihren Ausgang von der Idee nehmen, dass Ursachen einen Unterschied für ihre Wirkungen machen. Hierbei ist es dann gleichgültig, ob dieses 'Unterschiedmachen' über lokale kausale Prozesse vermittelt ist oder nicht. Entsprechend werden dann auch Fälle von negativer Kausalität und Kausalität durch Doppelverhinderung als genuine Fälle von Kausalität klassifiziert. Solche Ansätze können als *Abhängigkeitstheorien der Kausalität* bezeichnet werden.

Drittens könnte man aber versuchen, diese beiden Ansätze zu versöhnen. Dies ist jedoch zumindest auf den ersten Blick nicht leicht zu leisten: Jede (vollständige und einheitliche) Theorie der Kausalität wird die Fälle von negativer Verursachung und Doppelverhinderung entweder als Fälle von genuiner Kausalität klassifizieren und die Lokalitätsbedingung sowie die Idee einer physischen Verknüpfung und des Leistens kausaler Arbeit – zumindest in ihren allgemeinsten Formen – aufgeben *oder* das Gegenteil tun und die Fälle von negativer Kausalität und Doppelverhinderung nicht als Fälle von genuiner Kausalität klassifizieren. Es ist jedoch nicht möglich, an Produktion als der einzigen Art von genuiner Kausalität festzuhalten und dennoch Fälle von negativer Kausalität und Kausalität durch Doppelverhinderung zuzulassen. Ebenso ist es nicht möglich, Fälle von negativer Kausalität und Kausalität durch Doppelverhinderung als Fälle genuiner Kausalität zu akzeptieren und dennoch eine allgemeine Lokalitätsbedingung sowie eine allgemeine Verknüpfung zwischen Kausalität und kausaler Arbeit zu vertreten.[17]

Eine vierte Option, auf die dargestellte Lage zu reagieren, besteht schließlich in einem Pluralismus: Während die Lokalitätsbedingung, die Idee der physischen

[17] Einen interessanten Ansatz zur Versöhnung bietet jedoch Hüttemann (2013a: 191–192) und Hüttemann (2020: 1504–1505): Die Grundidee ist dabei, dass Ursachen als Störungen von Inertialprozessen verstanden werden können. Bei dem Umgang mit Fällen von negativer Kausalität und Doppelverhinderung kommt dann viel darauf an, wie die relevanten Inertialprozesse bestimmt werden. Insofern der Prozess, der darin besteht, dass die Anwesenheit des Stoppschilds den Unfall verhindert, als Inertialprozess aufgefasst wird, kann Susis Diebstahl als Störung dieses Inertialprozesses und somit als Ursache verstanden werden. Der Diebstahl steht dann durchaus in einem raumzeitlichen Kontakt zu diesem Inertialprozess. Allerdings ist der relevante Inertialprozess selbst teilweise über negative Ereignisse bestimmt. Eine kurze Diskussion der Lokalitätsbedingung in dieser Störungstheorie der Kausalität findet sich in Hüttemann (2013a: 206–207).

Verknüpfung und die Idee der kausalen Arbeit in Bezug auf einen Begriff der Kausalität – kausale Produktion – durchaus ihre Berechtigung haben, reicht es für die Anwendbarkeit eines anderen Begriffs von Kausalität – kausale Abhängigkeit – bereits aus, wenn ein Faktor einen Unterschied für einen anderen Faktor macht – unabhängig von der Frage, ob es eine lokale Verknüpfung zwischen den Faktoren gibt.

Wie immer man sich mit Blick auf diese vier Optionen entscheidet, ein Punkt sollte relativ unkontrovers sein: Kausale Produktion und kausale Abhängigkeit sind distinkte Relationen, die in Fällen von negativer Kausalität und Doppelverhinderung auseinanderfallen. Man kann die beiden Relationen begrifflich trennen. Allein dieser Punkt reicht als Ausgangspunkt für die weitere Diskussion schon aus. Ich konzentriere mich dabei im Folgenden auf die Relation der kausalen Abhängigkeit und spreche dabei mitunter schlicht von ‘Kausalität’. Die Unterscheidung zwischen kausaler Abhängigkeit und kausaler Produktion wird aber in Abschnitt 4.5. wieder relevant.

4.1.2 Kausale Abhängigkeit und hinreichende Verursachung

In typischen Formulierungen von Exklusionsargumenten wird an einigen Stellen vom Begriff der hinreichenden Verursachung Gebrauch gemacht. Insbesondere beziehen sich die üblichen Formulierungen der These der kausalen Geschlossenheit (siehe 3.2.4.) und des Exklusionsprinzips auf diesen Begriff. Was hat es mit hinreichender Verursachung genauer auf sich?

In einer ersten Annäherung können hinreichende Ursachen als Ereignisse aufgefasst werden, die *für sich genommen nomologisch hinreichend* für ihre Wirkungen sind. Eine Möglichkeit, dies ein Stück weit zu präzisieren, ist die folgende:

> **Hinreichende Verursachung**: Ein Ereignis u ist genau dann eine hinreichende Ursache für ein Ereignis w, wenn gilt: Das Auftreten von u impliziert in Konjunktion mit den Naturgesetzen, dass w eintritt.[18]

Gegeben diese Bestimmung von hinreichender Verursachung lässt sich leicht erkennen, dass die uns aus dem Alltag bekannten Ursachen in aller Regel keine *hinreichenden* Ursachen sind. Mein Stoß gegen die weiße Kugel beim Billard ist beispielsweise keine *hinreichende* Ursache dafür, dass die 11er Kugel in die

[18] Ein Verweis auf Ereignistypen erübrigt sich hier soweit ich sehe unter der Voraussetzung, dass Ereignisse als Eigenschaftsinstantiierungen aufzufassen sind.

Tasche fällt. Denn das Auftreten meines Stoßes allein impliziert *nicht* in Konjunktion mit den Naturgesetzen, dass die 11er Kugel in die Tasche fällt. Vielmehr müssen die Umstände günstig sein. Beispielsweise führt mein Stoß nur dann dazu, dass die 11er Kugel in die Tasche fällt, wenn mein Mitspieler nicht auf den Tisch greift, um die Kugel aufzuhalten. Auch würde mein Stoß nicht dazu führen, dass die Kugel in die Tasche fällt, wenn unmittelbar nach meinem Stoß der Tisch zusammenbrechen oder ein Komet einschlagen würde. Für sich genommen ist der Stoß also nicht hinreichend. Nur unter der Annahme, dass keine ungünstigen Umstände in der Form von bereitstehenden Verhinderern gegeben sind, impliziert das Auftreten des Stoßes in Konjunktion mit den Naturgesetzen, dass die Kugel in die Tasche fällt.

Hinreichend kann eine Ursache also erst sein, wenn solche möglichen Verhinderungen durch das Auftreten der Ursache naturgesetzlich ausgeschlossen sind. Und das ist plausiblerweise erst dann der Fall, wenn eine ausreichend große Raumzeitregion betrachtet wird. Betrachten wir beispielsweise die weitere Umgebung des Billardtischs und stellen dabei fest, dass kein Mitspieler bereitsteht, um die Kugel aufzuhalten, ist dieser potentielle Verhinderer schonmal ausgeschlossen. Nicht ausgeschlossen aber wären weniger alltägliche Verhinderer wie ein Komet, der sich mit großer Geschwindigkeit aus großer Entfernung dem Billardtisch nähert. Um auch solche potentiellen Verhinderer auszuschließen, muss die betrachtete Raumzeitregion ausgeweitet werden. Eine sicher ausreichend große Raumzeitregion haben wir erst, wenn wir einen Gesamtausschnitt des Vergangenheitslichtkegels der Wirkung zum Zeitpunkt der Ursache betrachten. Denn erst durch die Spezifikation einer solchen – gigantischen – Raumzeitregion ist ausgeschlossen, dass potentielle Verhinderer mit (bis zu) Lichtgeschwindigkeit unterwegs sind, um die 11er Kugel aufzuhalten.[19]

Ein Ereignis, das einen Gesamtausschnitt des Vergangenheitslichtkegels der Wirkung umfasst, ist sicherlich nicht die Art von Ereignis, die wir im Alltag als Ursache ansprechen. Der Begriff der hinreichenden Verursachung weicht in diesem Sinne in seiner Extension von einem alltäglichen Verständnis von Verursachung ab: Nur sehr komplexe und weiträumige Ereignisse sind plausible

[19] Vgl. Loewer (2007: 252): "It is a commonplace among philosophers of science, but perhaps not as recognized as it should be outside of philosophy of science, that for any small region R of space at time t nothing much short of the state of the universe in a sphere with center R and whose radius is one light second (i.e. 186,000 miles) at t – 1 second is causally sufficient for determining what will occur (or the chances at t – 1 of what will occur) in R."

Kandidaten für hinreichende Ursachen. Klarerweise sind gewöhnliche Ursachen im Sinne von kausaler Abhängigkeit nicht hinreichend für ihre Wirkungen.[20]

Auch wenn gewöhnliche Ursachen aber keine *im strengen Sinne* hinreichenden Ursachen sind, könnten sie noch immer *in den gegebenen Umständen* hinreichend sein. Eine Annäherung an die Idee der hinreichenden Verursachung *in den Umständen* kann wie folgt gegeben werden:

> **Hinreichende Verursachung in den Umständen**: Ein Ereignis u ist genau dann eine *in den Umständen* hinreichende Ursache für ein Ereignis w, wenn gilt: Das Auftreten von u impliziert in Konjunktion mit den Naturgesetzen *und den aktualen Umständen*, dass w eintritt.

Gewöhnliche lokale Ereignisse wie mein Billard-Stoß können *in den Umständen* hinreichend für ihre Wirkungen sein. Das Problem, dass im strengen Sinne hinreichende Ursachen sehr komplex sein müssen, wandelt sich jetzt zu der Einsicht, dass eine vollständige Spezifikation der Umstände sehr komplex sein muss. Damit mein Billardstoß gemeinsam mit den Umständen und den Naturgesetzen impliziert, dass die weiße Kugel in die Tasche fällt, müssen die ‚Umstände' eine Spezifikation eines Ausschnitts des Vergangenheitslichtkegels der Wirkung zum Zeitpunkt des Stoßes (minus den Stoß) umfassen. Denn nur so können sämtliche

---

[20] Eine andere Frage ist es, ob hinreichende Ursachen als Ursachen im Sinne von kausaler Abhängigkeit verstanden werden können. Macht das Auftreten des komplexen, ausgedehnten Ereignisses, das einen vollständigen Ausschnitt aus dem Vergangenheitslichtkegel der Wirkung (dass die 11er Kugel in die Tasche fällt) umfasst, einen Unterschied für das Auftreten der Wirkung? Dies scheint nicht unbedingt gegeben. Denn es gibt zahlreiche Weisen, wie dieses komplexe Ereignis nicht stattfinden kann. Einige dieser Weisen würden dazu führen, dass die Wirkung nicht stattfindet. Beispielsweise besteht eine Art, wie das komplexe Ereignis nicht stattfinden kann, darin, dass ich die weiße Kugel nicht stoße – während sich alles weitere nach Möglichkeit so verhält, wie tatsächlich. Unter diesen Umständen würde auch die 11er Kugel nicht in die Tasche fallen.

Das komplexe Ereignis enthält jedoch auch viele irrelevante ‚Teile'. Zum Beispiel ist in der Spezifikation des Ausschnitts des Vergangenheitslichtkegels der Wirkung zum Zeitpunkt meines Stoßes auch spezifiziert, dass Susi 20 km entfernt ihre Hand hebt. Eine andere Weise, wie das komplexe Ereignis nicht stattfinden kann, ist daher, dass Susi 20 km entfernt nicht ihre Hand hebt – während sich alles weitere nach Möglichkeit so verhält, wie tatsächlich. Bei dieser Alternative würde ich stoßen und die Kugel würde in die Tasche fallen. Es gibt keinen Grund, diese Weise, wie das komplexe Ereignis nicht eintritt, für fernliegender zu halten als die Weise, in der ich nicht stoße. Es gilt daher nicht: Wenn das komplexe Ereignis nicht stattfinden würde, dann würde die 11er Kugel nicht in die Tasche fallen.

Diese Argumentation beruht jedoch auf einer Ersetzungsevaluation des relevanten kontrafaktischen Konditionals. Unter Voraussetzung einer Entfernungs-Evaluation wäre das relevante Konditional wohl wahr. Diese Begriffe werden in 4.3.2 eingeführt.

mögliche Verhinderer ausgeschlossen werden. Es stellt sich nun die Frage, ob Abhängigkeitsverursachung mit so verstandener *in den Umständen hinreichender Verursachung* gleichgesetzt werden kann.

Hiergegen argumentieren Vertreter*innen von Abhängigkeitstheorien der Kausalität vehement. So schreibt etwa Peter Menzies:

> The most obvious meaning that philosophers have in mind when they talk about sufficient causes is that sufficient causes are events instantiating properties that are nomologically sufficient for the properties instantiated by the effect events. More precisely, an F-event is a sufficient cause of a G-event just in case the property F, in conjunction with laws of nature and other properties instantiated in the background conditions, logically implies property G. Such a conception fits with a deductive-nomological conception of causation, which often lies in the background of many discussions of mental causation. But this conception's influence is completely unwarranted and philosophers of causation have long lamented its distorting effects.[21]

Menzies Verständnis von (in den Umständen) hinreichender Verursachung entspricht dem oben eingeführten. Er kritisiert diesen Ansatz unter anderem dafür, dass er zahlreiche irrelevante Faktoren als in den Umständen hinreichende Ursachen klassifiziert. Wesley Salmon führt unter anderem die folgende Geschichte zur Verdeutlichung dieses Irrelevanz-Problems an:

> The moon reappeared after a lunar eclipse, for the people made a great deal of noise, banging on pots and pans and setting off fireworks, and the moon always reappears after an eclipse when much noise occurs.[22]

Der Lärm ist hier offenbar irrelevant für das Wiedererscheinen des Mondes. Dennoch folgt daraus, dass der Lärm auftaucht und den aktualen Umständen in Konjunktion mit den Naturgesetzen, dass der Mond wiedererscheint. Also ist der Lärm nach Maßgabe der obigen Definition eine in den Umständen hinreichende Ursache. Gleichzeitig ist der Lärm keine Ursache im Sinne von kausaler Abhängigkeit: Der Mond wäre auch dann wiedererschienen, wenn die Leute keinen Lärm gemacht hätten. Menzies folgert aus diesem Problem, dass in den Umständen hinreichende Verursachung kein genuiner kausaler Begriff ist.[23]

---

[21] Menzies (2015: 27).

[22] Salmon (1971: 34). Menzies (2015: 27–28) zitiert ein anderes Beispiel aus diesem Aufsatz.

[23] Ähnliche Kritik findet sich bei Raatikainen (2010: 360), Raatikainen (2018: 33–34), Woodward (2003: 194) sowie Woodward (2008a).

Eine andere Reaktion könnte darin bestehen, die oben eingeführten groben Charakterisierungen hinreichender Verursachung zu modifizieren oder zu ersetzen. Dies würde den Weg zu modernen Regularitätstheorien der Kausalität weisen.[24] Ich halte es für eine offene Frage, ob sich ausgehend von der Idee, dass Ursachen in den Umständen hinreichende Faktoren sind, eine Theorie der Kausalität entwickeln lässt, die die Grundidee, dass Ursachen einen Unterschied für ihre Wirkungen machen, angemessen einfängt. Da Regularitätstheorien jedoch nicht im Mittelpunkt dieser Arbeit stehen, gehe ich hierauf nicht näher ein.

Wichtig ist jedoch, dass es einen Unterschied zwischen im strengen Sinne hinreichenden Ursachen und Ursachen im Sinne von kausaler Abhängigkeit gibt. Auf die Rolle von im strengen Sinne hinreichender Verursachung in den Exklusionsargumenten gehe ich in Abschnitt 4.2. noch einmal ein.

4.1.3 Kausale Abhängigkeit und Proportionalität

Einige Autor*innen gehen davon aus, dass (Abhängigkeits-)Ursachen proportional zu ihren Wirkungen sein müssen: Um als genuine Ursachen zu zählen, dürfen sie weder zu viel unnötiges Detail enthalten bzw. zu spezifisch sein, noch dürfen sie nötiges Detail auslassen bzw. zu unspezifisch sein.[25] Diese Grundidee lässt sich zunächst an einem vielzitierten Beispiel von Stephen Yablo verdeutlichen:

> **Yablos Taube**: Die Taube Sophie hat gelernt, auf alle roten und nur auf rote Objekte zu picken. Gerade legt jemand ein scharlachrotes Objekt vor Sophie und Sophie beginnt, zu picken.[26]

Dass jemand ein rotes Objekt vor Sophie gelegt hat, ist die proportionale Ursache dafür, dass Sophie pickt. Zwei eng mit diesem Ereignis zusammenhängende Ereignisse sind jedoch nicht proportional zu Sophies Verhalten: Das Ereignis, das darin besteht, dass jemand ein *farbiges* Objekt vor Sophie abgelegt hat, ist beispielsweise zu unspezifisch. Nicht jedes farbige Objekt hätte dazu geführt, dass

[24] Paul und Hall (2013: 14–16) skizzieren einen Ansatz, von der Idee hinreichender Verursachung ausgehend eine Definition von Kausalität zu entwickeln. Baumgartner (2013a) verteidigt einen neueren regularitätstheoretischen Ansatz. Weslake (2017: 225–227) entwickelt eine Definition hinreichender Verursachung im Rahmen eines interventionistischen Ansatzes.

[25] Vgl. Yablo (1992: 274): „[Causes] should incorporate a good deal of causally important material but not too much that is causally unimportant".

[26] Yablo (1992: 257).

Sophie pickt. Dass das Objekt rot ist, ist also eine wesentliche Information für die Erklärung von Sophies Verhalten und darf deshalb nicht unspezifiziert bleiben. Das Ereignis, das darin besteht, dass jemand ein *scharlachrotes* Objekt vor Susi abgelegt hat, ist hingegen zu spezifisch, um als proportionale Ursache zu zählen: Auch ein bordeauxrotes Objekt (oder ein Objekt in irgendeinem anderen Rotton) hätte dazu geführt, dass Sophie pickt. Dass das Objekt diesen spezifischen Rotton hatte, ist also irrelevant für Sophies Verhalten.

Dieses Beispiel beruht – wie auch Yablos eigene Explikation von Proportionalität[27] – auf einem Vergleich zwischen Ereignissen, deren konstitutive Eigenschaften in einer Determinationsrelation (im Sinne von Abschnitt 2.5.4.) stehen. Die Grundidee proportionaler Verursachung ist jedoch nicht auf Anwendungen beschränkt, bei denen Determination eine Rolle spielt.[28] Daher ist es vielleicht sinnvoll, die Idee noch an einem anderen Beispiel (von Hilary Putnam) zu verdeutlichen:

> Suppose we have a very simple physical system – a board in which there are two holes, a circle one inch in diameter and a square one inch high, and a cubicle peg one-sixteenth of an inch less than one inch high. We have the following very simple fact to explain: *the peg passes through the square hole, and it does not pass through the round hole.*[29]

Putnam diskutiert zwei mögliche Erklärungen dafür, dass das kubische Objekt zwar durch die quadratische Öffnung passt, nicht aber durch die runde Öffnung. Die erste – übermäßig detaillierte und daher nicht besonders gute – Erklärung betrachtet das kubische Objekt sowie das Brett mit den Öffnungen als Atomgitter und berechnet davon ausgehend alle möglichen Trajektorien der beiden Systeme. Die Erklärung besteht dann darin, dass es mögliche Trajektorien des kubischen Objekts gibt, bei denen es durch die quadratische Öffnung durchgeht, aber keine möglichen Trajektorien, bei denen es durch die runde Öffnung durchgeht. Die zweite mögliche Erklärung, die deutlich naheliegender ist, beruft sich schlicht auf die geometrischen Eigenschaften der involvierten Objekte: Das kubische Objekt passt durch die quadratische Öffnung, weil sie groß genug ist. Es passt nicht durch die runde Öffnung, weil sie zu klein ist.

Man kann dieses Beispiel so umdeuten, dass hier Kausalbeziehungen eine Rolle spielen: Die proportionale Ursache dafür, dass das kubische Objekt durch die quadratische Öffnung passt, besteht in seiner geometrischen Form, nicht in

[27] Vgl. Yablo (1992: 273–279).

[28] Vgl. z. B. Bontly (2005: 335–336) und List und Menzies (2009: 480–481).

[29] Putnam (1975: 130).

seiner genauen mikrophysikalischen Struktur. Wenn es eine andere mikrophysikalische Struktur gehabt hätte aber Größe und Form beibehalten hätte, hätte es noch immer durch die Öffnung gepasst. Die mikrophysikalische Struktur des Objekts wäre als Ursache zu spezifisch. Auch hier hat die Grundidee der proportionalen Verursachung also eine einleuchtende Anwendung. Zugleich ist man hierdurch nicht darauf festgelegt, dass die Mikrostruktur des Objekts eine *Determinate* der geometrischen Eigenschaften des Objekts ist.

Im Allgemeinen kommt die Idee der proportionalen Verursachung aber in Kontexten zur Anwendung, in denen mehrere Kandidaten für Ursachen in einer Relation der ontologischen Abhängigkeit stehen. Die Röte des Objekts bei Yablos Taube ist ontologisch abhängig von der Scharlachröte des Objekts. Die geometrische Form des Objekts in Putnams Beispiel ist ontologisch abhängig von der Mikrostruktur des Objekts. In beiden Fällen wird dabei ein Ereignis als proportionale Ursache ausgeschlossen, das *fundamentaler* ist als die proportionale Ursache. Die Scharlachröte des Objekts in Yablos Taube ist fundamentaler als die Röte, weil die Röte von der Scharlachröte abhängig ist. Die geometrische Form des Objekts in Putnams Beispiel ist fundamentaler als die Mikrostruktur des Objekts, weil die geometrische Form von der Mikrostruktur abhängig ist. Dies verweist bereits auf eine mögliche Anwendung im Zusammenhang mit mentaler Verursachung im nicht-reduktiven Physikalismus. Es ist durchaus möglich, dass höherstufige Ereignisse wie mentale Ereignisse proportional zu bestimmten Wirkungen sind, während die physischen Ereignisse, von denen die mentalen Ereignisse abhängen, nicht proportional zu denselben Wirkungen sind, weil sie etwa zu spezifisch sind. Hier wird also bereits deutlich, dass Überlegungen zur Proportionalität von Ursachen zu einer Zurückweisung der Prinzipien der Verursachung im Stufenmodell der Realität führen könnten. Die Idee der Proportionalität kann entsprechend gerade für inkompatibilistsische Erwiderungen auf Exklusionsargumente wichtige argumentative Arbeit leisten.

Ein weiteres wichtiges Merkmal von proportionaler Verursachung betrifft ihr Verhältnis zur im vorigen Abschnitt eingeführten Relation der hinreichenden Verursachung: Hinreichende Ursachen sind nicht unbedingt proportional zu ihren Wirkungen.[30] Sie können insbesondere leicht zu viel unnötiges Detail enthalten. Dass ein scharlachrotes Objekt vor Sophie gelegt wird, ist (in den Umständen) hinreichend dafür, dass Sophie pickt. Wie wir gesehen haben ist dieses Ereignis aber eben *nicht* proportional zu Sophies Picken.

[30] Der Punkt gilt sowohl für im strengen Sinne hinreichende Verursachung als auch für in den Umständen hinreichende Verursachung.

Es gibt verschiedene Ansätze, die Grundidee proportionaler Verursachung genauer zu fassen. In den Abschnitten 4.3.12. und 4.4.10. gehe ich hierauf noch einmal ein. An dieser Stelle sollte die vorläufige Erläuterung durch die beiden Beispiele aber reichen.

Wie verhält sich die Idee der proportionalen Verursachung zu Kausalität als kausaler Abhängigkeit? Hierzu gehen die Meinungen auseinander: Während einige Autor*innen eine Proportionalitätsbedingung in ihre Definition kausaler Abhängigkeit einbauen und entsprechend ausschließlich proportionale Ursachen als Abhängigkeits-Ursachen klassifizieren[31], meinen andere, dass die Relation kausaler Abhängigkeit unabhängig von einer Proportionalitätsforderung verstanden werden sollte.[32]

Zunächst scheint es naheliegend, zumindest eine enge Verwandtschaft zwischen der Idee der proportionalen Verursachung und der Idee der kausalen Abhängigkeit zu erkennen: Proportionale Ursachen sind in besonderem Maße dazu geeignet, als Abhängigkeitsursachen aufgefasst zu werden. In Yablos Taube macht die Röte des Objekts den ausschlaggebenden Unterschied, nicht die Scharlachröte oder die Farbigkeit. Und der Grund hierfür hat mit für kausale Abhängigkeit typischen kontrafaktischen Überlegungen zu tun: Wäre das Objekt nicht rot gewesen, hätte Sophie nicht gepickt. Aber wäre das Objekt nicht scharlachrot (sondern bordeauxrot) gewesen, hätte Sophie noch immer gepickt.[33]

Auf der anderen Seite zeigt sich, dass die Proportionalitätsforderung als allgemeine Anforderung an kausale Abhängigkeit vielleicht zu stark ist. Denn zahlreiche Ereignisse, die wir intuitiv klarerweise als Ursachen auffassen würden, sind nicht proportional zu ihren Wirkungen. Beispielsweise ist es klarerweise eine Ursache für Sokrates Tod, dass er ein Getränk mit Schierling getrunken hat. Auch scheint zu gelten, dass Sokrates Trinken aus dem Schierlingsbecher einen Unterschied für seinen Tod gemacht hat: Hätte er nicht getrunken, wäre er nicht gestorben. Jedoch ist dieser Ereignis nicht proportional zu Sokrates Tod. Denn auch wenn Sokrates ein anderes Gift getrunken hätte, wäre er gestorben. In diesem Sinne ist das Ereignis des Trinkens eines *Schierlingsgetränks* zu spezifisch. Die proportionale Ursache für Sokrates Tod ist das Trinken von Gift. Bei genauerem Hinsehen mag aber auch dieses Ereignis noch zu spezifisch sein. Denn für den Tod hätte es auch gereicht, wenn Sokrates etwas anderes

[31] Vgl. Yablo (1992), List und Menzies (2009), Harbecke (2014), Zhong (2014), Pernu (2016).

[32] Vgl. z. B. Kroedel (2019) für eine Diskussion kausaler Abhängigkeit im Kontext mentaler Verursachung, die keine Proportionalitätsforderung berücksichtigt.

[33] Eine gründlichere Diskussion dieser Behauptungen findet sich in Abschnitt 4.3.12.

tödliches getan hätte. Vielleicht käme also erst ‘etwas tödliches tun’ als proportionale Ursache in Frage.[34] Wenn Proportionalität also als notwendige Bedingung für kausale Abhängigkeit verstanden wird, ist kausale Abhängigkeit ein weitaus rareres Phänomen, als gedacht.

Es ist also problematisch, Proportionalität als notwendige Bedingung für kausale Abhängigkeit aufzufassen. Hieraus folgt jedoch nicht, dass die Idee proportionaler Verursachung keine wichtige Arbeit in einer Theorie der Kausalität und einer kausalistischen Erwiderung auf den Vorwurf des Epiphänomenalismus leisten kann. Denn es gibt zwei verschiedene Weisen, wie eine Proportionalitätsforderung in eine Theorie der Kausalität eingebaut werden kann: Erstens kann Proportionalität eine notwendige Bedingung für Verursachung sein. In diesem Falle kann man von einer *starken Proportionalitätsforderung* sprechen. Proportionalität diskriminiert dann zwischen Ursachen und Ereignissen, die keine Ursachen sind. Starke Proportionalität hat die soeben geschilderten Probleme. Zweitens kann Proportionalität aber auch als ein Merkmal verstanden werden, das bestimmte Ursachen vor anderen Ursachen auszeichnet. Es gibt dann proportionale Ursachen und nicht-proportionale Ursachen. Es bleibt – auch vor dem Hintergrund der geschilderten Probleme von starker Proportionalität – plausibel, dass proportionale Ursachen gewissermaßen *bessere* Ursachen sind als nicht-proportionale Ursachen. Eine solche *schwache Proportionalitätsforderung* scheint also durchaus attraktiv.

4.2 Kausalität, hinreichende Verursachung und Exklusionsargumente

In diesem Abschnitt gehe ich auf die Rolle der Relation hinreichender Verursachung im Vorwurf des Epiphänomenalismus ein. In Abschnitt 4.2.1. geht es um die Frage, ob die Annahme der Existenz mentaler Verursachung im Sinne hinreichender Verursachung verstanden werden sollte. Abschnitte 4.2.2. und 4.2.3. gehen genauer auf die Rolle hinreichender Verursachung im einfachen Exklusionsargument und damit in der These der kausalen Geschlossenheit und dem üblichen Exklusionsprinzip ein. Abschnitt 4.2.4. schließlich diskutiert die Rolle hinreichender Verursachung in den Erweiterungen des einfachen Exklusionsarguments und somit in den drei Prinzipien der Verursachung im Stufenmodell der Realität.

[34] Vgl. Bontly (2005: 339–342) und McDonnell (2017: 1468–1470).

4.2.1 Hinreichende Verursachung und mentale Verursachung

Der Vorwurf des Epiphänomenalismus etabliert in einem ersten Schritt, dass aus dem nicht-reduktiven Physikalismus folgt, dass es keine mentale Verursachung gibt. In einem zweiten Schritt wird dann behauptet, dass es mentale Verursachung gibt. Dann aber muss der nicht-reduktive Physikalismus falsch sein. Gibt es eine starke Version des Vorwurfs des Epiphänomenalismus, die sich in beiden Schritten auf hinreichende Verursachung bezieht?

Nach dem in 3.1.2. Gesagten sind im strengen Sinne hinreichende Ursachen sehr komplexe und weiträumige Ereignisse. Klarerweise kommen mentale Ereignisse nicht in diesem Sinn als hinreichende Ursachen in Frage. Mein Wunsch, die 11er Kugel in die Tasche zu stoßen, ist eben so wenig im strengen Sinne hinreichend für meinen Stoß gegen die weiße Kugel, wie mein Stoß im strengen Sinne hinreichend dafür ist, dass die 11er Kugel in die Tasche fällt. Denn mein Wunsch führt nur unter günstigen Umständen zu meinem Stoß. Wenn mein Mitspieler mich zum Beispiel kurz vor meinem Stoß wehschubst, stoße ich nicht, obwohl ich möchte. Eine hinreichende Ursache müsste diesen möglichen Verhinderer ausschließen. Und mein Wunsch allein vermag dies nicht.

Die Annahme, dass mentale Ereignisse im strengen Sinne hinreichende Ursachen für ihre Wirkungen sind, hat also – unabhängig vom nicht-reduktiven Physikalismus – schon gar keine hohe Ausgangsplausibilität. Nicht-reduktive Physikalist*innen sollten sich daher nicht davon beeindrucken lassen, wenn ihnen nachgewiesen werden kann, dass sich diese Annahme aus ihrer Position ergibt. Entsprechend sollte sich der zweite Schritt des Vorwurfs des Epiphänomenalismus *nicht* auf im strengen Sinne hinreichende Verursachung beziehen. Vielmehr sollte er sich auf Verursachung simpliciter beziehen.

4.2.2 Hinreichende Verursachung, kausale Geschlossenheit und das übliche Exklusionsprinzip

Schon in Abschnitt 3.2.1. habe ich dafür argumentiert, dass die These der kausalen Geschlossenheit sich auf den Begriff der hinreichenden Verursachung beziehen muss. Grund hierfür war, dass Strukturen gemeinsamer Verursachung mit nicht-physischen Teilursachen nach allgemeinem Verständnis gegen die These der kausalen Geschlossenheit sprechen.[35] Wenn es physische Wirkungen gibt,

[35] Vgl. Abbildung 3.1.

die nur in Zusammenarbeit von einer physischen Teilursache und einer mentalen (oder magischen) Teilursache zustande kommen, ist die These der kausalen Geschlossenheit nach allgemeinem Verständnis falsch. Um solche Situationen aber auszuschließen, muss sich die These der kausalen Geschlossenheit auf im strengen Sinne hinreichende Verursachung beziehen. Würde die These der kausalen Geschlossenheit nämlich bloß verlangen, dass jedes physische Ereignis eine physische Ursache simpliciter (oder eine in den Umständen hinreichende physische Ursache) hat, würde sie solche Kausalstrukturen nicht ausschließen.

Auch das übliche Exklusionsprinzip beruht auf dem Begriff der hinreichenden Verursachung. Dies hat zwei Gründe:

Erstens muss das übliche Exklusionsprinzip zur These der kausalen Geschlossenheit ‚passen', um eine Funktion im einfachen Exklusionsargument zu übernehmen. Die These der kausalen Geschlossenheit liefert wie gerade gesehen den Input, dass jedes physische Ereignis (das überhaupt eine hinreichende Ursache hat), eine *hinreichende* physische Ursache hat. Exklusionsprinzipien müssen daran anknüpfen und sich daher im Antezedens auf hinreichende Verursachung beziehen: Wenn ein Ereignis eine *hinreichende* Ursache hat, dann hat es darüber hinaus keine weitere Ursache – es sei denn, es handelt sich um einen Fall genuiner Überdetermination.

Zweitens werden Exklusionsprinzipien nur dann plausibel, wenn sie sich im Antezedens auf hinreichende Verursachung beziehen. Denn es ist unstrittig, dass ein Ereignis mehrere Ursachen zu einem Zeitpunkt t haben kann, ohne überdeterminiert zu sein. Ein Brand kann sowohl von einem Kurzschluss als auch von der (gleichzeitigen) Anwesenheit von brennbarem Material verursacht werden. Tatsächlich sind in diesem Beispiel beide Ursachen notwendig, um den Brand zu verursachen. Jegliche Fälle von gemeinsamer Verursachung sind Gegenbeispiele gegen die Annahme, dass ein Ereignis nur eine Ursache haben kann, wenn es nicht genuin überdeterminiert ist.[36] Das übliche Exklusionsprinzip sollte aber nicht so interpretiert werden, dass es die offensichtlich falsche Aussage trifft, dass jedes nicht genuin überdeterminierte Ereignis nur eine Ursache hat. Das Antezedens des Exklusionsprinzips muss also den Begriff der hinreichenden Verursachung verwenden.[37]

---

[36] Vgl. Árnadóttir and Crane (2013: 257 f.) und Kroedel (2019: 157).

[37] Auch kann es sich nicht auf den Begriff der in den Umständen hinreichenden Ursache berufen, da die Umstände weitere Ursachen enthalten können. Der Kurzschluss ist in den Umständen hinreichend. Dennoch gibt es weitere Ursachen, wie die Anwesenheit von brennbarem Material.

Das Konsequenz des Exklusionsprinzips hingegen kann, aber muss sich nicht auf den Begriff der hinreichenden Verursachung beziehen. Es gibt zwei prima facie plausible Versionen des üblichen Exklusionsprinzips:

> **Das übliche Exklusionsprinzip mit Verursachung im Konsequenz**: Wenn ein Ereignis eine hinreichende Ursache hat, dann hat es darüber hinaus keine weitere *Ursache* – es sei denn, es handelt sich um einen Fall von genuiner Überdetermination.
>
> **Das übliche Exklusionsprinzip mit hinreichender Verursachung im Konsequenz**: Wenn ein Ereignis eine hinreichende Ursache hat, hat es darüber hinaus keine weitere *hinreichende Ursache* – es sei denn, es handelt sich um einen Fall von genuiner Überdetermination.

Um das übliche Exklusionsprinzip mit Verursachung im Konsequenz jedoch zu plausibilisieren, muss der Begriff der genuinen Überdetermination auch Fälle umfassen, in denen ein Ereignis eine hinreichende Ursache und darüber hinaus eine weitere *Ursache* hat. Im Fall der zwei Attentäter ist es plausibel, dass beide Schüsse *hinreichende* Ursachen sind. Was aber, wenn einer der Schüsse nicht für sich genommen zum Tod des Opfers geführt hätte?

Wechseln wir zur Veranschaulichung zu einem anderen Beispiel: In einem Fluss stirbt ein Fisch, weil zwei verschiedene Firmen giftiges Abwasser in den Fluss geleitet haben. Für den Tod des Fisches ist es notwendig und hinreichend, dass 100 l Abwasser in den Fluss geleitet werden. Firma A hat exakt 100 l Abwasser in den Fluss geleitet. Firma B hat unabhängig davon und gleichzeitig zusätzlich 50 l Abwasser in den Fluss geleitet. Es scheint, dass der Tod des Fisches in einem gewissen Sinne überdeterminiert ist: Er hat eine hinreichende Ursache und darüber hinaus eine weitere *Ursache*. Denn es scheint plausibel,

auch die Handlung von Firma B als Ursache für den Tod des Fisches zu zählen.[38] Jedoch handelt es sich eben nicht um eine *hinreichende* Ursache.[39]

Der Begriff der genuinen Überdetermination muss auch solche Fälle unter sich fassen, damit das übliche Exklusionsprinzip mit Verursachung im Konsequenz plausibel wird. Denn anderenfalls würde die Möglichkeit solcher Fälle gegen das übliche Exklusionsprinzip mit Verursachung im Konsequenz sprechen.

Setzt man voraus, dass genuine Überdetermination dadurch ausgezeichnet ist, dass die überdeterminierenden Ursachen ontologisch unabhängig voneinander sind, sind solche Fälle von Überdetermination mit einer hinreichenden Ursache und einer zusätzlichen unabhängigen Ursache aber ohnehin schon inbegriffen. Denn die Handlung von Firma A und die Handlung von Firma B stehen in keiner Beziehung der ontologischen Abhängigkeit. Der Unterschied zu der abhängigen Überdetermination, die viele nicht-reduktive Physikalist*innen bei mentaler Verursachung am Werk sehen, ist also bewahrt. Das übliche Exklusionsprinzip mit Verursachung im Konsequenz würde also nicht durch den geschilderten Fall in Verlegenheit gebracht.

Das übliche Exklusionsprinzip mit Verursachung im Konsequenz und das übliche Exklusionsprinzip mit hinreichender Verursachung im Konsequenz stehen, soweit ich sehe, mit Blick auf ihre Plausibilität auf einer Stufe. Die in Abschnitt 3.3.10. diskutierten Begründungen lassen sich auf beide Versionen anwenden. Wenn bereits eine hinreichende Ursache vorliegt, dann kann eine zusätzliche Ursache aus Einfachheitsgründen ebenso weggekürzt werden wie eine zusätzliche hinreichende Ursache. Und wenn kausale Arbeit nicht ‚doppelt' verrichtet werden kann (wie im Fall zweier hinreichender Ursachen), dann kann sie plausiblerweise auch nicht ‚anderthalbfach' verrichtet werden (wie im Fall einer hinreichenden Ursache und einer zusätzlichen Teilursache).

---

[38] Wer die Intuition nicht teilt, lässt sich vielleicht hiervon überzeugen:

1) Wenn Firma A nur 99 l Abwasser in den Fluss geleitet hätte, wäre die Handlung von Firma A nicht mehr hinreichend für den Tod des Fisches gewesen. Der Fall wäre dann ein Fall gemeinsamer Verursachung. Es ist klar, dass die Handlung von Firma B in diesem Fall als Ursache zu zählen ist. Wieso sollte die Handlung von Firma B aber aufhören, Ursache zu sein, wenn Firma B 1 l mehr Abwasser in den Fluss leitet? Wer die Handlung von Firma B also als Ursache klassifiziert, wenn Firma A 99 l Abwasser einleitet, sollte sie auch als Ursache klassifizieren, wenn Firma A 100 l einleitet.

2) Wenn Firma B keine Ursache für den Tod des Fisches ist, und man nur für Handlungskonsequenzen verantwortlich sein kann, die man verursacht hat, dann trägt Firma B auch keine Verantwortung für den Tod des Fisches. Es scheint aber unplausibel, dass Firma B sich moralisch aus der Affäre ziehen kann, weil sie das Glück hatte, dass Firma A statt 99 l 100 l Abwasser in den Fluss geleitet hat.

[39] Ähnliche Beispiele finden sich u. a. bei Ferey und Dehez (2016: 657).

Jedoch stehen die beiden Prinzipien in ihrer argumentativen Funktionalität keinesfalls auf einer Stufe. Denn auf Grundlage des üblichen Exklusionsprinzips mit *hinreichender* Verursachung im Konsequenz lässt sich im Rahmen von Exklusionsargumenten nur dafür argumentieren, dass mentale Ereignisse nicht *hinreichend* für ihre Wirkungen sind. Es lässt sich nicht etablieren, dass sie überhaupt keine Wirkungen haben. Man betrachte etwa die folgende Variante des einfachen Exklusionsarguments:

Das einfache Exklusionsargument gegen hinreichende mental-physische Verursachung

Es sei m ein mentales Ereignis. p und p* seien zwei physische Ereignisse. Unter diesen Annahmen lässt sich das einfache Exklusionsargument wie folgt rekonstruieren:

(1) p ist hinreichende Ursache für p* (motiviert durch die These der kausalen Geschlossenheit).

(2) m ist nicht mit p (oder einem Teil von p) identisch (motiviert durch die These der Nicht-Identität).

(3) Wenn p hinreichende Ursache von p* ist und m nicht mit p (oder einem Teil von p) identisch ist und p* nicht genuin überdeterminiert ist, dann ist m nicht *hinreichende* Ursache von p* (motiviert durch das übliche Exklusionsprinzip mit hinreichender Verursachung im Konsequenz).

(4) p* ist nicht genuin überdeterminiert.

(5) Also: m ist nicht *hinreichende* Ursache von p* (aus 1, 2, 3, 4).

Das übliche Exklusionsprinzip mit hinreichender Verursachung im Konsequenz motiviert nur die Annahme in Zeile (3): Wenn p hinreichende Ursache von p* ist und m nicht mit p (oder einem Teil von p) identisch ist und p* nicht genuin überdeterminiert ist, dann ist m nicht *hinreichende* Ursache von p*. Deshalb erlaubt sie auch nur die Konklusion (5), dass m nicht *hinreichende* Ursache für p* ist.

Wie in 4.2.1. argumentiert, ist dies aber eine Konsequenz, von der sich nicht-reduktive Physikalist*innen nicht beeindrucken lassen sollten. Exklusionsargumente sollten, um den nicht-reduktiven Physikalismus in ernsthafte Schwierigkeiten zu bringen, etablieren, dass es keine mentale Verursachung gibt und nicht bloß, dass es keine hinreichende mentale Verursachung gibt.

4.2.3 Kritik am Begriff hinreichender Verursachung und Varianten des Exklusionsprinzips

Einige Autor*innen sehen den Begriff der hinreichenden Verursachung – wie in 3.2.4 und 4.1.2. schon angesprochen – kritisch. Sie meinen, dass es sich nicht um einen genuinen kausalen Begriff handelt. Vor diesem Hintergrund schreibt etwa Panu Raatikainen über die These der kausalen Geschlossenheit und das übliche Exklusionsprinzip:

> Both these assumptions involve confusing causes with sufficient conditions. There are causes, which are difference-makers; and there are sufficient conditions, which are wholly different issues and not causes of any sort; there are no such things as sufficient causes. Hence, I do not think that these two assumptions are so much as false (or true) as mongrels based on a conceptual confusion which fail to make clear sense. After all, the whole point of the exclusion argument and the debate surrounding it is to ask whether the mental is capable of being a cause of something physical. But then, surely the argument and its premises should talk about causes and not be formulated in terms of sufficient conditions.[40]

Für Raatikainen ist der Ausdruck 'hinreichende Verursachung' also eine Fehlbezeichnung. Hinreichende Bedingungen sind keine Ursachen und Ursachen keine hinreichenden Bedingungen. Eine solche Einstellung kann unter anderem durch eine starke Proportionalitätsforderung an Kausalität motiviert werden. Hier kommen hinreichende Bedingungen (typischerweise) nicht als Ursachen in Frage, weil sie die Proportionalitätsforderung nicht erfüllen. Sollte Raatikainen damit recht haben, dass die These der kausalen Geschlossenheit und das übliche Exklusionsprinzip aufgrund ihrer Verwendung des Begriffs hinreichender Verursachung sinnlos sind, wäre das natürlich ein großes Problem für Vertreter*innen von Exklusionsargumenten.

Man kann auf zumindest zweierlei Weise hierauf reagieren:

Erstens kann man die Sinnhaftigkeit des Begriffs der hinreichenden Verursachung verteidigen. Tatsächlich scheint es – wenn eine starke Proportionalitätsforderung abgelehnt wird – nicht unbedingt problematisch, bestimmte hinreichende Bedingungen als Ursachen anzusprechen. Auch wenn man Raatikainen aber zugesteht, dass der Begriff hinreichender Verursachung sinnlos ist, kann man an der Schlagkräftigkeit von Exklusionsargumenten festhalten:

[40] Raatikainen (2010: 360).

Zweitens kann man die These der kausalen Geschlossenheit und das übliche Exklusionsprinzip nämlich auf eine Weise modifizieren, die den Verweis auf hinreichende Verursachung vermeidet. Dabei muss man Raatikainens Kommentare berücksichtigen, dass "the whole point of the exclusion argument is to ask whether the mental is capable of being a cause" und dass "surely the argument and its premise should talk about causes". Hier gibt es zumindest zwei Möglichkeiten:

Erstens muss das Argument – wie in 4.2.1. betont – in erster Linie in seiner *Konklusion* über Verursachung simpliciter sprechen. Die *Prämissen* müssen hingegen nicht durchgehend nur auf Verursachung simpliciter bezogen sein, sondern lediglich so beschaffen sein, dass sie eine Konklusion über Verursachung simpliciter erlauben. Will man den Begriff der hinreichenden Verursachung vermeiden, kann man stattdessen einfach den Begriff der nomologisch hinreichenden Bedingung verwenden. Die These der kausalen Geschlossenheit wird im einfachen Exklusionsargument dann durch die These der nomologischen Geschlossenheit ersetzt (siehe 3.2.5.). Das übliche Exklusionsprinzip kann durch das folgende nomologische Exklusionsprinzip ersetzt werden:

> **Nomologisches Exklusionsprinzip**: Wenn es für ein Ereignis eine nomologisch hinreichende Bedingung zu t gibt, dann hat es darüber hinaus keine weitere Ursache zu t – es sei denn, es handelt sich um einen Fall von genuiner Überdetermination.

Auf Basis dieses nomologischen Exklusionsprinzips lässt sich ebenfalls eine Variante des einfachen Exklusionsarguments formulieren, das etabliert, dass mentale Ereignisse keine physischen Wirkungen haben.

Allerdings übertragen sich zumindest nicht alle Begründungen des üblichen Exklusionsprinzips auch auf das nomologische Exklusionsprinzip. Mit Blick auf das nomologische Exklusionsprinzip ist es nicht einmal prima facie plausibel, eine Begründung auf der Idee kausaler Arbeit aufzubauen. Ein Begründungsversuch über Einfachheitsüberlegungen ist hier schon besser übertragbar: Wenn bereits ein physisches Ereignis nomologisch hinreichend für eine physische Wirkung ist, können wir uns die Annahme einer zusätzlichen mentalen Ursache sparen. Letztlich konnte diese Begründung aber schon in Bezug auf das übliche Exklusionsprinzip nicht überzeugen (siehe 3.4.3.). In der finalen Einschätzung ist das nomologische Exklusionsprinzip daher genauso gut (d. h. schlecht) begründet wie das übliche Exklusionsprinzip.

Wichtig mit Blick auf das nomologische Exklusionsprinzip ist folgendes: Selbst nicht-reduktive Physikalist*innen, die den Begriff hinreichender Verursachung für sinnlos halten, sind auf eine Ablehnung des nomologischen

Exklusionsprinzips festgelegt. Eine inkompatibilistische Erwiderung auf Exklusionsargumente (3.4.5.) kann daher nicht auf eine Akzeptanz des nomologischen Exklusionsprinzips hinauslaufen. Insofern Inkompatibilist*innen also ein Exklusionsprinzip akzeptieren, ist dies *nicht* das nomologische Exklusionsprinzip.

Die zweite Möglichkeit, wie man die These der kausalen Geschlossenheit und das übliche Exklusionsprinzip in Reaktion auf die Kritik am Begriff hinreichender Verursachung modifizieren kann, besteht darin, die beiden Prinzipien durchgehend auf Verursachung simpliciter zu beziehen. Auf den ersten Blick mag dies wenig aussichtsreich wirken. Denn wie geschildert gibt es gute Gründe, die Prinzipien unter Bezug auf hinreichende Verursachung zu formulieren.

Mit Blick auf die These der kausalen Geschlossenheit schließt dieser gute Grund – dass eine These der kausalen Geschlossenheit, die sich auf Verursachung simpliciter bezieht, zu schwach wäre, um Strukturen gemeinsamer Verursachung mit nicht-physischen Teilursachen auszuschließen – allerdings nicht die *Wahrheit* der entsprechenden These der kausalen Geschlossenheit in Bezug auf Verursachung simpliciter aus. Insofern schon die Prämisse, dass jedes physische Ereignis, das überhaupt eine Ursache hat, auch eine physische Ursache hat, eine Rolle in einem Exklusionsargument übernehmen kann, spricht zunächst einmal nichts dagegen, sich auf diese Prämisse zu beziehen.

In Bezug auf das übliche Exklusionsprinzip habe ich in 4.2.2. argumentiert, dass eine Version, die sich auf Verursachung simpliciter bezieht, nicht plausibel ist: Denn es gibt unstrittigerweise Fälle von gemeinsamer Verursachung, die keine Fälle genuiner Überdetermination sind. Ereignisse haben regelmäßig mehr als eine Ursache zu einem Zeitpunkt. Man kann jedoch ein anderes Exklusionsprinzip vertreten, das sich ebenfalls nur auf Verursachung simpliciter beruft und einen Verweis auf hinreichende Verursachung vermeidet. Und zwar ist das eine Variante des in 3.3.9. schon kurz angerissenen Exklusionsprinzips der ontologischen Abhängigkeit

> **Das Exklusionsprinzip der ontologischen Abhängigkeit**: Wenn ein Ereignis u eine Ursache für ein Ereignis w ist und ein Ereignis h von u ontologisch abhängt, dann ist Ereignis h keine Ursache für Ereignis w.[41]

[41] Vgl. z. B. List und Menzies (2009) für Diskussion. List und Menzies argumentieren, dass das Prinzip zwar nicht allgemein wahr ist, jedoch in typischen Fällen mentaler Verursachung Anwendung findet und zum Ausschluss einer Ursache führt. Aufgrund einer starken Proportionalitätsforderung meinen sie jedoch, dass in typischen Fällen mentaler Verursachung nicht das mentale Ereignis als Ursache ausgeschlossen wird, sondern die physische Basis des mentalen Ereignisses.

Das Exklusionsprinzip der ontologischen Abhängigkeit findet nur auf Situationen Anwendung, in denen mehrere vermeintliche Ursachen in einer Relation der ontologischen Abhängigkeit stehen. Es kann eventuell vor dem Hintergrund bestimmter Theorien der Kausalität motiviert werden.[42] In Kombination mit der Version der These der kausalen Geschlossenheit, die sich auf Verursachung simpliciter bezieht, ermöglicht dieses Prinzip ein Exklusionsargument, das sich durchgehend auf Verursachung simpliciter bezieht:

Das einfache Exklusionsargument der ontologischen Abhängigkeit
Es sei m ein mentales Ereignis. p und p* seien zwei physische Ereignisse. Unter diesen Annahmen lässt sich das einfache Exklusionsargument wie folgt rekonstruieren:

(1) p ist eine Ursache für p* (motiviert durch die entsprechend modifizierte These der kausalen Geschlossenheit).

(2) m ist ontologisch abhängig von p (motiviert durch die These der ontologischen Abhängigkeit).

(3) Wenn p eine Ursache für p* ist und m von p ontologisch abhängig ist, dann ist m keine Ursache für p* (motiviert durch das Exklusionsprinzip der ontologischen Abhängigkeit).

(4) Also: m ist keine Ursache von p* (aus 1, 2, 3).

Auch diese Variante eines Exklusionsarguments vermeidet einen Bezug auf hinreichende Verursachung. Auf den Status des ausschlaggebenden Exklusionsprinzips der ontologischen Abhängigkeit in der kontrafaktischen Theorie und der interventionistischen Theorie der Kausalität komme ich in den Abschnitten 4.3.11. und 4.4.6. – 4.4.10. noch zurück.

[42] So diskutiere ich in 4.4.8. und 4.4.9. einen Versuch, ein entsprechendes Prinzip im Rahmen einer interventionistischen Theorie zu begründen.

4.2.4 Hinreichende Verursachung und die drei Prinzipien der Verursachung im Stufenmodell der Realität

Bis hierher beschränkt sich meine Diskussion hinreichender Verursachung auf das einfache Exklusionsargument. Dieses Argument ist jedoch – wie in 3.3.2. argumentiert – in seiner Reichweite eingeschränkt: Es kann bestenfalls etablieren, dass mentale Ereignisse im nicht-reduktiven Physikalismus keine $\text{physischen}_{\text{eng}}$ Wirkungen haben. Um es zu einem allgemeinen Argument gegen mentale Verursachung auszubauen, sollte man sich auf eines der in Abschnitt 3.3.5. eingeführten Prinzipien der Verursachung im Stufenmodell der Realität berufen. Wie verhält sich der Begriff der hinreichenden Verursachung zu diesen drei Prinzipien? Ich gehe die drei Prinzipien der Reihe nach knapp durch:

Zum Prinzip der Abwärtsverursachung:

> **Prinzip der Abwärtsverursachung**: Wenn ein Ereignis u eine (hinreichende) Ursache für ein Ereignis w ist und Ereignis w ontologisch abhängig von einem Ereignis b ist, dann ist u auch eine (hinreichende) Ursache für Ereignis b.

Die Funktion dieses Prinzips innerhalb des Arguments aus dem Prinzip der Abwärtsverursachung besteht darin, das folgende Konditional zu etablieren: Wenn ein mentales Ereignis m ein höherstufiges Ereignis h verursacht, dann verursacht m auch eine $\text{physische}_{\text{eng}}$ Basis p* von h. Das einfache Exklusionsargument etabliert dann, dass m nicht p* verursacht. Per *modus tollens* wird geschlossen, dass m nicht h verursacht.

Um diese Funktion zu erfüllen, muss das Prinzip sich (zumindest im Antezedens)[43] auf die Relation der *Verursachung* beziehen. Denn die Anwendung des *modus tollens* führt zu der Falschheit des Antezedens des Konditionals. Wenn das Antezedens also in einer Aussage über *hinreichende* Verursachung besteht, dann kommt auch nur die Verneinung einer Aussage über *hinreichende* Verursachung

[43] Das Prinzip dürfte sich, was seine argumentative Funktion angeht, im Konsequenz auch auf hinreichende Verursachung beziehen. Das Argument aus dem Prinzip der Abwärtsverursachung würde sich im zweiten Schritt dann auf das in diesem Abschnitt rekonstruierte einfache Exklusionsargument gegen *hinreichende* mental-physische Verursachung berufen. Jedoch hat ein solches gemischtes Prinzip, was seine Plausibilität anbelangt, bedeutende Nachteile gegenüber dem unten angeführten Prinzip: Wieso sollte aus dem Bestehen einer Verursachungsrelation etwas für das Bestehen einer hinreichenden Verursachungsrelation folgen?

Zudem würde das Argument dann noch immer nicht etablieren, dass mentale Ereignisse keine $\textit{physischen}_{\textit{eng}}$ Wirkungen haben, sondern nur, dass sie nicht hinreichend für diese Wirkungen sind.

als Konklusion des Arguments in Frage. Wir haben aber schon gesehen, dass die gewünschte Konklusion die Verneinung einer Aussage über *Verursachung* sein sollte. Also muss sich das Prinzip der Abwärtsverursachung auf *Verursachung* und nicht auf *hinreichende Verursachung* beziehen. Die Version des Prinzips, die die gewünschte Rolle im Argument einnimmt, ist also diese:

> **Prinzip der Abwärtsverursachung – V**: Wenn ein Ereignis u eine Ursache für ein Ereignis w ist und Ereignis w ontologisch abhängig von einem Ereignis b ist, dann ist u auch eine Ursache für Ereignis b.

Die in Abschnitt 3.3.5. genannte Motivation für dieses Prinzip überträgt sich auf den ersten Blick auf diese Version des Prinzips: Wenn Susis Wunsch einen Unterschied dafür macht, ob sie ihre Hand hebt und ihr Handheben von einer physischen Basis ontologisch abhängt, dann macht ihr Wunsch scheinbar auch einen Unterschied dafür, ob die zugrundeliegende physische Basis eintritt. Denn in jeder Welt, in der Susi ihre Hand nicht hebt, liegt auch die physische Basis des Handhebens nicht vor. Tatsächlich wird sich zeigen, dass sich das Prinzip auch im Rahmen einer kontrafaktischen Theorie der Kausalität (Abschnitt 4.3.11.) und einer interventionistischen Theorie der Kausalität (Abschnitt 4.4.9.) begründen lässt.

Zum Prinzip der Aufwärtsverursachung:

> **Prinzip der Aufwärtsverursachung**: Wenn ein Ereignis u eine (hinreichende) Ursache für ein Ereignis w ist und ein Ereignis h ontologisch abhängig von Ereignis w ist, dann ist u auch eine (hinreichende) Ursache für Ereignis h.

Die Funktion dieses Prinzips innerhalb des Arguments aus dem Prinzip der Aufwärtskausalität besteht darin, zu etablieren, dass nicht nur physische$_{eng}$ Ereignisse hinreichende physische$_{eng}$ Ursachen haben, sondern auch alle höherstufigen Ereignisse hinreichende physische$_{eng}$ Ursachen haben. Da die physischen$_{eng}$ Basen aufgrund der These der kausalen Geschlossenheit hinreichende physische$_{eng}$ Ursachen haben und die Relation der hinreichenden Verursachung die Stufenleiter hinaufklettert, haben auch höherstufige Ereignisse hinreichende physische$_{eng}$ Ursachen. Das Exklusionsprinzip wird dann direkt auf höherstufige (vermeintliche) Wirkungen mentaler Ereignisse angewandt: Diese haben bereits hinreichende physische$_{eng}$ Ursachen, so dass mentale Ursachen ausgeschlossen werden.

Um diese Funktion zu erfüllen, muss sich das Prinzip auf *hinreichende* Verursachung beziehen. Es wird – ähnlich wie das Exklusionsprinzip im einfachen

Exklusionsargument – von der These der kausalen Geschlossenheit ‚gefüttert‘: Die These der kausalen Geschlossenheit liefert den Input, dass die $\text{physische}_{\text{eng}}$ Basis p* vom höherstufigen Ereignis h eine hinreichende $\text{physische}_{\text{eng}}$ Ursache hat. Das Prinzip der Aufwärtskausalität muss daran anknüpfen und etablieren, dass auch h diese hinreichende $\text{physische}_{\text{eng}}$ Ursache p hat. Das Antezedens muss sich daher auf hinreichende Verursachung beziehen, weil die These der kausalen Geschlossenheit sich auf hinreichende Verursachung bezieht.

Zugleich ‚füttert‘ das Prinzip der Aufwärtskausalität in diesem Argument das Exklusionsprinzip. Der Output ist, dass auch das höherstufige Ereignis h eine hinreichende Ursache p hat. Das Exklusionsprinzip knüpft daran an und etabliert, dass h dann keine weiteren Ursachen hat. Auch das Konsequenz muss sich daher auf hinreichende Verursachung beziehen, da das Antezedens des Exklusionsprinzips sich auf hinreichende Verursachung beruft.

Unabhängig von diesen funktionalen Überlegungen wurde das Prinzip der Aufwärtsverursachung in Abschnitt 3.3.5. auch auf eine Weise motiviert, die nahelegt, dass es sich auf hinreichende Verursachung bezieht: Wenn p garantiert, dass p* eintritt und p* garantiert, dass h eintritt, dann garantiert p auch, dass h eintritt. Diese Motivation wird nur dann plausibel, wenn es in dem Prinzip um Behauptungen über hinreichende Verursachung geht.

Für das Argument aus dem Prinzip der Aufwärtsverursachung ist also die folgende Version ausschlaggebend:

> **Prinzip der Aufwärtsverursachung – HV**: Wenn ein Ereignis u eine *hinreichende Ursache* für ein Ereignis w ist und ein Ereignis h ontologisch abhängig von Ereignis w ist, dann ist u auch eine *hinreichende Ursache* für Ereignis h.

Diese Bemerkungen beschränken sich jedoch auf die Verwendung des Prinzips innerhalb des in 3.3.7. eingeführten Arguments aus dem Prinzip der Aufwärtskausalität. Tatsächlich wird es in späteren Teilen der Arbeit – insbesondere in der Diskussion kontrafaktischer Theorien der Kausalität – auch noch einmal um die Version gehen, die sich auf Verursachung und nicht auf hinreichende Verursachung bezieht. Denn in Kombination mit dem Exklusionsprinzip der ontologischen Abhängigkeit kann dieses Prinzip ebenfalls in einem Exklusionsargument verwendet werden.

Zum Prinzip der basalen Verursachung:

> **Prinzip der basalen Verursachung**: Wenn ein Ereignis u eine (hinreichende) Ursache für ein Ereignis w ist und Ereignis u ontologisch abhängig von einem Ereignis b ist, dann ist auch Ereignis b eine (hinreichende) Ursache für Ereignis w.

Die Funktion dieses Prinzips innerhalb des Arguments aus dem Prinzip der basalen Verursachung besteht darin, das folgende Konditional zu etablieren: *Wenn ein mentales Ereignis m irgendein Ereignis w verursacht, dann gibt es eine weitere, physische*$_{eng}$ *und hinreichende Ursache für w – nämlich p, die Basis von m.* Das Exklusionsargument wird dann auf die von der Annahme, dass m w verursacht, implizierte Situation angewandt, in der m w verursacht und die Basis p eine hinreichende Ursache für w ist. Diese wird als unmöglich verworfen. Denn *wenn p eine hinreichende Ursache für w ist, dann ist m eben keine Ursache für w*.

Um diese Funktion zu erfüllen, muss das Prinzip ein Antezedens haben, dass sich auf Verursachung bezieht. Das ist nötig, um zu garantieren, dass sich die Konklusion des Arguments auf Verursachung bezieht. Es muss aber ein Konsequenz haben, das sich auf hinreichende Verursachung bezieht. Denn es ‚füttert' das Exklusionsprinzip. Was also verlangt wäre, wäre die folgende Version:

> **Das gewünschte Prinzip der basalen Verursachung**: Wenn ein Ereignis u eine *Ursache* für ein Ereignis w ist und Ereignis u ontologisch abhängig von einem Ereignis b ist, dann ist Ereignis b eine *hinreichende* Ursache für Ereignis w.

Aus der Behauptung, dass ein höherstufiges Ereignis ein anderes Ereignis *verursacht*, können wir nach diesem Prinzip ableiten, dass die Basis des höherstufigen Ereignisses eine *hinreichende* Ursache ist.

Unglücklicherweise passt die Motivation, die ich in Abschnitt 3.3.5. für das Prinzip der basalen Verursachung gegeben habe, nicht sehr gut zu dieser Version. Denn diese berief sich auf die folgende Überlegung: Wenn b eine Basis von u ist, dann garantiert b, dass u eintritt. Wenn u aber eine hinreichende Ursache für w ist, dann garantiert u, dass w eintritt. Also garantiert auch b (über u), dass w eintritt.

Diese Motivation setzt voraus, dass u eine hinreichende Ursache für w ist. So kann also allenfalls die folgende Version begründet werden:

> **Das begründete Prinzip der basalen Verursachung**: Wenn ein Ereignis u eine *hinreichende Ursache* für ein Ereignis w ist und Ereignis u ontologisch abhängig von einem Ereignis b ist, dann ist Ereignis b eine hinreichende Ursache für Ereignis w.

Diese Version wiederum kann nicht in einem Argument verwendet werden, das etabliert, dass die Annahme, dass m w *verursacht*, selbstwiderlegend ist. Dieses Prinzip kann stattdessen ‚lediglich' in einem Argument verwendet werden, das etabliert, dass die Annahme, dass m eine *hinreichende* Ursache für w ist, selbstwiderlegend ist.

Was ist also von dieser Einsicht zu halten? Ist das Argument aus dem Prinzip der basalen Verursachung nun wertlos? Drei Punkte:

Erstens könnte das gewünschte Prinzip der basalen Verursachung doch noch durch andere Überlegungen gestützt werden. Wenn – wie Karen Bennett argumentiert[44] – die Umstände, die eine Kernbasis zu einer totalen Basis machen, zugleich die Umstände sind, die eine Ursache zu einer hinreichenden Ursache machen, könnte es durchaus plausibel sein, dass das mentale Ereignis eine Ursache ist, während die totale Basis eine hinreichende Ursache ist. Zwar ist bezogen auf die Kernbasis von Susis Wunsch zu erwarten, dass sie – ebenso wie Susis Wunsch selbst – nur eine Ursache simpliciter für Susis Armbewegung ist. Um aus der Kernbasis aber eine totale Basis zu machen, müssen wir zahlreiche Umstände hinzufügen. Wir erhalten dann ein komplexes physisches Ereignis, das Susis Wunsch necessitiert. Karen Bennett meint nun, dass eben jene Umstände, die benötigt werden, um aus der Kernbasis des Wunsches eine totale Basis zu machen, auch benötigt werden, um zu einer hinreichenden Ursache für Susis Armheben zu gelangen. Das würde bedeuten, dass zumindest die totale Basis von Susis Wunsch eine hinreichende Ursache für ihr Armheben ist. Entsprechend könnte auf diese Weise ein Prinzip der basalen Verursachung verteidigt werden, das die gewünschte argumentative Rolle im Argument aus dem Prinzip der basalen Verursachung zu spielen vermag.

Zweitens könnte bereits die Konklusion, die sich auf Grundlage des plausiblen Prinzips der basalen Verursachung – HV etablieren lässt, interessant genug sein. Die Annahme, dass ein mentales Ereignis eine hinreichende Ursache für ein anderes Ereignis ist, wäre dann selbstwiderlegend. Dies bleibt zumindest eine überraschende Einsicht.

Drittens kann man ein Prinzip der basalen Verursachung, das sich sowohl im Antezedens als auch im Konsequenz auf Verursachung simpliciter bezieht („Wenn ein Ereignis u eine *Ursache* für ein Ereignis w ist und Ereignis u ontologisch abhängig von einem Ereignis b ist, dann ist auch Ereignis b eine *Ursache* für Ereignis w.“) mit dem Exklusionsprinzip der ontologischen Abhängigkeit („Wenn ein Ereignis u eine Ursache für Ereignis w ist und Ereignis b von u ontologisch abhängt, dann ist u keine Ursache für w“) kombinieren, um ein Exklusionsargument für die gewünschte Konklusion zu erhalten, dass die Annahme mentaler Verursachung selbstwiderlegend ist. Die Version des Prinzips der basalen Verursachung, die sich sowohl im Antezedens als auch im Konsequenz auf Verursachung simpliciter bezieht, wird daher ebenfalls noch einmal Thema im Kontext der kontrafaktischen und interventionistischen Theorie der Kausalität.

[44] Bennett (2003: 488–489).

Insgesamt lässt sich also folgendes Resümee zur Rolle hinreichender Verursachung in den Erweiterungen des einfachen Exlusionsarguments ziehen: Das Prinzip der Aufwärtsverursachung und das Prinzip der Abwärtsverursachung können scheinbar auf der Grundlage der schon in Abschnitt 3.3.5. gegebenen vorläufigen Motivation in der Version verteidigt werden, die für das Funktionieren ihrer in 3.3.6. bis 3.3.8. eingeführten zugehörigen Exklusionsargumente nötig ist. Beim Prinzip der basalen Verursachung gestaltet sich die Sache etwas schwieriger, aber auch hier könnte eine Begründung in der gewünschten Version vielversprechend sein. Alle drei Prinzipien können zudem auf eine Weise formuliert werden, die durchgehend von einem Verweis auf hinreichende Verursachung frei ist und sich ausschließlich auf Verursachung simpliciter bezieht. Diese Versionen der drei Prinzipien können in Kombination mit dem in 4.2.3. erwähnten Exklusionsprinzip der ontologischen Abhängigkeit in weiteren Exklusionsargumenten vorkommen. Im Kontext der kontrafaktischen Theorie und der interventionistischen Theorie der Kausalität werden diese Versionen der drei Prinzipien daher weiter diskutiert.

4.3 Die kontrafaktische Theorie der Kausalität und mentale Verursachung im nicht-reduktiven Physikalismus

Kontrafaktische Theorien der Kausalität knüpfen recht unmittelbar an die in Abschnitt 4.1. erläuterte Grundidee an, dass Ursachen Faktoren sind, die einen Unterschied für ihre Wirkungen machen. Sie gehen dabei davon aus, dass dieses Unterschiedmachen auf der Grundlage kontrafaktischer Konditionale expliziert werden kann.

In diesem Unterkapitel diskutiere ich die Anwendung dieser paradigmatischen Abhängigkeitstheorie der Kausalität auf mentale Verursachung im nicht-reduktiven Physikalismus. Dabei verfahre ich wie folgt: In den Abschnitten 4.3.1. bis 4.3.4. führe ich die kontrafaktische Theorie der Kausalität zunächst unabhängig von der Frage nach mentaler Verursachung im nicht-reduktiven Physikalismus ein. Abschnitt 4.3.1. stellt die einfachst mögliche Version einer kontrafaktischen Theorie der Kausalität dar. Abschnitt 4.3.2. diskutiert die Semantik kontrafaktischer Konditionale. Abschnitt 4.3.3. führt zwei Komplikationen bei der semantischen Bewertung kontrafaktischer Konditionale ein, die im Zusammenhang mit der Anwendung der kontrafaktischen Theorie der Kausalität auf mentale Verursachung im nicht-reduktiven Physikalismus eine Rolle spielen. In

Abschnitt 4.3.4. diskutiere ich knapp zwei typische Einwände gegen die einfache kontrafaktische Theorie der Kausalität.

Die verbleibenden Abschnitte diskutieren darauf aufbauend die Anwendung der kontrafaktischen Theorie der Kausalität auf mentale Verursachung im nicht-reduktiven Physikalismus. In Abschnitt 4.3.5. führe ich zunächst ein Schema eines einfachen Arguments für mentale Verursachung im nicht-reduktiven Physikalismus ein. In den Abschnitten 4.3.6. bis 4.3.9. diskutiere ich vier Argumente für die zentrale Prämisse dieses Arguments, die besagt, dass physische Ereignisse von mentalen Ereignissen kontrafaktisch abhängen. In Abschnitt 4.3.10 ziehe ich ein knappes Zwischenfazit. In Abschnitt 4.3.11. diskutiere ich den Status der drei Prinzipien der Verursachung im Stufenmodell der Realität. In den Abschnitten 4.3.12. bis 4.3.14. diskutiere ich schließlich die Einbindung einer Proportionalitätsforderung im Rahmen der kontrafaktischen Theorie und ihre Anwendung auf mentale Verursachung.

4.3.1 Die einfache kontrafaktische Theorie der Kausalität

Die Grundidee der einfachen kontrafaktischen Theorie der Kausalität kann wie folgt auf den Punkt gebracht werden: Kontrafaktische Abhängigkeit zwischen gänzlich distinkten, tatsächlich stattfindenden Ereignissen ist notwendig und hinreichend für Kausalität. Den Kern der kontrafaktischen Theorie der Kausalität macht also die folgende Definition aus:

> **Definition kontrafaktischer Kausalität (DKK)**: Ein Ereignis u verursacht ein Ereignis w genau dann, wenn gilt: (dkk-i) u und w finden tatsächlich statt, (dkk-ii) u und w sind gänzlich distinkt und (dkk-iii) w ist kontrafaktisch abhängig von u.

Bedingung (dkk-i) in (DKK) macht deutlich, dass das Ziel der Definition *aktuale* Verursachung ist. Kausal verknüpft können nur Ereignisse sein, die tatsächlich stattfinden. Bedingung (dkk-ii) stellt sicher, dass nur Ereignisse kausal verknüpft sind, die ‚gänzlich distinkt' voneinander sind. Distinktheit schließt hierbei mindestens eine metaphysische Necessitationsbeziehung zwischen den beiden Ereignissen aus. Im nicht-reduktiven Physikalismus sind mentale und andere höherstufige Ereignisse daher nicht im relevanten Sinne gänzlich distinkt von ihren physischen Basen. Entsprechend sind sie auch nicht kausal verknüpft.

Bedingung (dkk-iii) macht den Kern der einfachen kontrafaktischen Theorie der Kausalität aus. Hier wird verlangt, dass Wirkungen von ihren Ursachen

kontrafaktisch abhängig sind. Was ist dabei mit kontrafaktischer Abhängigkeit gemeint?

Die Beziehung der kontrafaktischen Abhängigkeit zwischen zwei Ereignissen wird üblicherweise wie folgt bestimmt:

> **Definition kontrafaktischer Abhängigkeit (DKA)**: Ein Ereignis w ist kontrafaktisch abhängig von einem Ereignis u genau dann, wenn gilt: (dka-i) Wenn u eingetreten wäre, dann wäre w eingetreten und (dka-ii) wenn u nicht eingetreten wäre, dann wäre w nicht eingetreten.

Kontrafaktische Abhängigkeit ist also durch zwei kontrafaktische Konditionale bestimmt: Das in (dka-i) genannte Konditional verbindet das Eintreten von u mit dem Eintreten von w und das in (dka-ii) genannte Konditional verbindet das Nicht-Eintreten von u mit dem Nicht-Eintreten von w.

Unter der verbreiteten Annahme, dass kontrafaktische Konditionale mit einem wahren Antezedens und einem wahren Konsequens – ich spreche im Folgenden hierbei von faktischen kontrafaktischen Konditionalen[45] – trivial wahr sind, erübrigt sich im Kontext der Definition kontrafaktischer Kausalität ein Bezug auf das erste kontrafaktische Konditional, das das Eintreten der beiden Ereignisse betrifft. In Bedingung (dkk-i) wird bereits festgelegt, dass u und w eintreten. Unter der genannten Annahme über faktische kontrafaktische Konditionale folgt aus (dkk-i), dass w eintreten würde, wenn u eintreten würde. Dies legt also die folgende Definition von Kausalität nahe:

> **Einfache kontrafaktische Kausalität (EKK)**: Ein Ereignis u verursacht ein Ereignis w genau dann, wenn gilt: (ekk-i) u und w finden tatsächlich statt, (ekk-ii) u und w sind gänzlich distinkt und (ekk-iii) wenn u *nicht* eingetreten wäre, dann wäre auch w *nicht* eingetreten.

Wie wir später sehen werden, lehnen einige Autor*innen die oben genannte Annahme, dass faktische kontrafaktische Konditionale trivial wahr sind, ab. Die Rolle des faktischen kontrafaktischen Konditionals (dka-i) wird also noch einmal

[45] Die paradoxe Formulierung ‚faktische kontrafaktische Konditionale' könnte darauf hinweisen, dass die Bezeichnung ‚kontrafaktische Konditionale' nicht sehr glücklich gewählt ist. Einige Autor*innen sprechen daher von ‚subjunctive conditionals' (vgl. hierzu z. B. Lewis (1973a: 4), J. Bennett (2003)). Dies wiederum ist u. a. deshalb keine ideale Bezeichnung, weil erstens auch im Englischen entsprechende Konditionale ohne Verwendung des Subjunktivs formuliert werden können und zweitens einige Sprachen kein Subjunktiv kennen.

wichtig. An dieser Stelle können wir jedoch erst einmal bei der Definition (EKK) bleiben.

In paradigmatischen Fällen von Verursachung liegt auch laut der einfachen kontrafaktischen Theorie der Kausalität zufolge eine Verursachungsrelation vor: Nehmen wir etwa den Fall, dass Susi einen Stein auf eine Fensterscheibe wirft und die Scheibe deshalb zerbricht. Susis Wurf ist eine Ursache für das Zerbrechen der Scheibe. Die einfache kontrafaktische Theorie der Kausalität stimmt zu: Susi wirft und die Scheibe zerbricht, so dass (ekk-i) erfüllt ist. Susis Wurf und das Zerbrechen der Scheibe sind gänzlich distinkt, so dass (ekk-ii) erfüllt ist. Schließlich ist das Zerbrechen der Scheibe von Susis Wurf kontrafaktisch abhängig: Wenn Susi nicht geworfen hätte, dann wäre die Scheibe nicht zerbrochen (ekk-iii). Also ist Susis Wurf eine Ursache für das Zerbrechen der Scheibe.

Auch der in Abschnitt 4.1. eingeführte Fall von Doppelverhinderung zählt nach der kontrafaktischen Theorie als ein Fall von Kausalität: Wenn Susi das Stoppschild nicht geklaut hätte, wäre es nicht zum Unfall gekommen. Der Diebstahl ist also eine Ursache für den Unfall. Zudem kommen auch negative Ereignisse als Ursachen in Frage: Das Unterlassen meines Nachbarn, meine Blumen zu gießen, verursacht das Austrocknen der Blumen, denn wenn mein Nachbar die Blumen gegossen hätte, wären sie nicht ausgetrocknet. Hiermit zeigt sich die kontrafaktische Theorie als typische Abhängigkeitstheorie der Kausalität.

Während die einfache kontrafaktische Theorie der Kausalität mit dem paradigmatischen Fall von Verursachung (sowie dem Fall von Doppelverursachung) problemlos umgehen kann, werden wir sehen, dass sie in anderen Fällen größere Probleme hat. Bevor ich auf diese Problemfälle (Abschnitt 4.3.4.) eingehe, sollen die Grundlagen der kontrafaktischen Theorie jedoch noch etwas genauer erläutert werden. Hierfür gehe ich im folgenden Abschnitt auf die Semantik kontrafaktischer Konditionale ein.

4.3.2 Kontrafaktische Konditionale: Ähnlichkeit zwischen möglichen Welten

Zentral für ein genaueres Verständnis der kontrafaktischen Theorie der Kausalität ist die semantische Bewertung kontrafaktischer Konditionale. In diesem Abschnitt führe ich knapp die mögliche Welten Semantik kontrafaktischer Konditionale ein, die der kontrafaktischen Theorie der Kausalität zugrunde liegt. Zudem weise ich auf einige offene Fragen in Zusammenhang mit der semantischen Bewertung kontrafaktischer Konditionale hin, die im Verlauf der weiteren Diskussion eine Rolle spielen.

Die Wahrheitsbedingungen kontrafaktischer Konditionale können nach der in weiten Teilen der Debatte vorausgesetzten Lewis/Stalnaker-Semantik[46] wie folgt bestimmt werden:

> **Wahrheitsbedingungen kontrafaktischer Konditionale (WKK)**: Das kontrafaktische Konditional ‚A > B' (gelesen: ‚Wenn A der Fall wäre, dann wäre B der Fall') ist wahr genau dann, wenn gilt: (wkk-i) Es gibt keine möglichen Welten, in denen p der Fall ist oder (wkk-ii) es gibt eine Welt, in der p und q der Fall sind, und die der aktualen Welt ähnlicher ist als jede Welt, in der p der Fall ist, aber q nicht der Fall ist.

Bedingung (wkk-i) betrifft kontrafaktische Konditionale, deren Antezedens *unmöglich* ist. Sie legt fest, dass solche ‚kontramöglichen' Konditionale allgemein wahr sind.[47] Dieser Sonderfall muss uns jedoch (zunächst) nicht interessieren. Bedingung (wkk-ii) gibt die Wahrheitsbedingungen für kontrafaktische Konditionale mit *möglichem* Antezedens an. Sie ist die ausschlaggebende Bedingung für die weitere Diskussion.

Die Intuition hinter (wkk-ii) ist die folgende: Wenn die geringsten Abweichungen von der aktualen Situation, die mit der Wahrheit des Antezedens einhergehen, zu einer Wahrheit des Konsequenz führen, ist ein kontrafaktisches Konditional wahr. Man betrachtet zur Bewertung des kontrafaktischen Konditionals also nicht jede mögliche Situation, in der das Antezedens des Konditionals wahr ist. Stattdessen betrachtet man nur solche Situationen, die der aktualen Situation möglichst ähnlich sind. Will man – um David Lewis' Beispiel[48] zu verwenden – die Wahrheit des Konditionals ‚Wenn Kängurus keine Schwänze hätten, würden sie umfallen' bewerten, ist dafür nicht *jede* mögliche Welt relevant, in der Kängurus keine Schwänze haben. Stattdessen sind sondern nur *jene* möglichen Welten relevant, in denen Kängurus keine Schwänze haben *und die der aktualen Welt in allen anderen Hinsichten sehr ähneln*. Welten, in denen Kängurus keine Schwänze, aber dafür riesige Füße haben sind also ebenso irrelevant wie Welten, in denen Kängurus keine Schwänze haben und fliegen können. Stattdessen betrachten wir nur Welten, die in ihren Naturgesetzen und Einzelheiten der aktualen Welt soweit gleichen, wie es eben geht, wenn Kängurus keine Schwänze haben.

---

[46] Vgl. Stalnaker (1968) und Lewis (1973a).

[47] Diese Konvention – die Lewis (1973a: 24–26) nicht für uneingeschränkt intuitiv gedeckt, aber praktisch hält – wird in jüngerer Zeit verstärkt kritisiert. Einige Autor*innen ziehen daher eine Unterscheidung zwischen wahren und falschen kontramöglichen Konditionalen. Vgl. hierfür z. B. Abschnitt 2.5. in Berto und Jago (2018).

[48] Lewis (1973a: 1).

Die Wahrheitsbedingungen kontrafaktischer Konditionale setzen voraus, dass sich mögliche Welten nach ihrer Ähnlichkeit zur aktualen Welt ordnen lassen. Einer solchen Ordnung zugrunde liegt ein *Vergleich* zwischen zwei Welten mit Blick auf ihre Ähnlichkeit zur aktualen Welt. Die ausschlaggebenden, vergleichenden Ähnlichkeitsurteile haben also die folgende Form: Welt w1 ist der aktualen Welt a *ähnlicher als* Welt w2. Oder: Welt w1 ist der aktualen Welt a *weniger ähnlich als* Welt w2. Oder: Welt w1 ist der aktualen Welt a *genauso ähnlich wie* Welt w2.[49]

Es wird dabei vorausgesetzt, dass sich zwei beliebige Welten immer mit Blick auf ihre Ähnlichkeit zur aktualen Welt[50] vergleichen lassen. Dabei kann es sein, dass sich die beiden Welten mit Blick auf ihre Ähnlichkeit zur aktualen Welt gleichen. Die Ähnlichkeitsrelationen sind transitiv: Wenn w1 ähnlicher zu a ist als w2 und w2 ähnlicher zu a ist als w3, dann ist w1 ähnlicher zu a als w3. Die aktuale Welt a hat einen Sonderstatus in Bezug auf ihre Ähnlichkeit zu sich selbst: Es gibt keine anderen Welten, die der aktualen Welt ähnlicher sind, als die aktuale Welt selbst.[51] Es ergibt sich so eine Ordnung möglicher Welten nach ihrer Ähnlichkeit zur aktualen Welt.

Eine andere Art, dieser Ordnung Ausdruck zu verleihen, besteht in der Rede eines Systems von verschachtelten *Sphären* möglicher Welten. Die größte Sphäre umfasst dabei alle Welten unabhängig von ihrer Ähnlichkeit zur aktualen Welt. Von hier ausgehend kann man kleinere Sphären definieren, die zunehmend Welten ausschließen, die weniger ähnlich sind als die Welten in der jeweiligen Sphäre. Die kleinste Sphäre umfasst dann nur die aktuale Welt (und gegebenenfalls Welten, die der aktualen Welt so ähnlich sind wie sie selbst) und schließt alle weniger ähnlichen Welten aus.

Wenn wir die vereinfachende und in den vorliegenden Kontexten harmlose Annahme treffen, dass es für jede Proposition A eine Menge von der aktualen Welt *ähnlichsten* Welten gibt, in denen A wahr ist[52], und eine Welt, in der eine

---

[49] Lewis (1973a: 48) führt als grundlegend die Relation ‚w1 ist a *mindestens so ähnlich wie* w2' ein. Die anderen Relationen lassen sich unter Bezug auf diese Relation definieren.

[50] Aus Gründen der einfacheren Darstellung gehe ich hier immer von einem Vergleich zur aktualen Welt aus und beschränke die Diskussion auf die Wahrheitsbedingungen kontrafaktischer Konditionale in der aktualen Welt. Zudem gehe ich davon aus, dass alle Welten zugänglich sind, was einen expliziten Verweis auf eine Zugänglichkeitsrelation zwischen möglichen Welten erübrigt. Natürlich sind die Bestimmung der Wahrheitsbedingungen kontrafaktischer Konditionale so wie die Anmerkungen zur Ähnlichkeitsrelation eigentlich allgemeiner.

[51] Dies ist also eine schwache Zentrierungsforderung. Ich gehe auf den Unterschied zu starker Zentrierung unten noch ein.

[52] Dies wird oft die ‚Limit Assumption' genannt. Die Alternative wäre, dass es zu jeder p-Welt eine der aktualen Welt *ähnlichere* p-Welt gibt. In diesem Falle gäbe es keine Menge

Proposition A wahr ist, abkürzend eine ‚A-Welt' nennen, können wir folgende Vereinfachung für die Wahrheitsbedingungen kontrafaktischer Konditionale mit möglichem Antezedens einführen:

> **Vereinfachte Wahrheitsbedingungen kontrafaktischer Konditionale**: Das kontrafaktische Konditional ‚A > B' (gelesen: ‚Wenn A der Fall wäre, dann wäre B der Fall' ist wahr genau dann, wenn gilt: die der aktualen Welt ähnlichsten A-Welten sind B-Welten.

Ich setze in der folgenden Diskussion diese einfache Bestimmung der Wahrheitsbedingungen kontrafaktischer Konditionale voraus.

Für die semantische Bewertung kontrafaktischer Konditionale ausschlaggebend sind demzufolge vergleichende Urteile über die Ähnlichkeit zwischen möglichen Welten. Die Ähnlichkeitsrelation ist dabei bis hierher nur durch einige allgemeine formale Einschränkungen bestimmt. Konkrete Urteile über die Wahrheit kontrafaktischer Konditionale sind jedoch auf weitaus konkretere Ähnlichkeitsurteile angewiesen. Lewis schreibt hierzu Folgendes:

> This analysis (plus some simple observations about the formal character of comparative similarity) is about all that can be said in full generality about counterfactuals. While not devoid of testable content – it settles some questions of logic – it does little to predict the truth values of particular counterfactuals in particular contexts. The rest of the study of counterfactuals is not fully general. Analysis 2 is only a skeleton. It must be fleshed out with an account of the appropriate similarity relation, and this will differ from context to context.[53]

Die Semantik kontrafaktischer Konditionale ist durch die formalen Eigenschaften der Ähnlichkeitsrelation also noch nicht hinreichend spezifiziert. Die formalen Eigenschaften der Ähnlichkeitsrelation schränken zwar die Logik kontrafaktischer Konditionale ein – es lassen sich Urteile über die (Un-)Gültigkeit einiger Schlussregeln für kontrafaktische Konditionale ableiten – jedoch erlauben sie keine direkten Urteile über die Wahrheitswerte konkreter kontrafaktischer Konditionale. ‘Wenn Kängurus keine Schwänze hätten, würden sie umfallen’ ist genau dann wahr, wenn Kängurus in den ähnlichsten Welten, in denen sie keine Schwänze haben, umfallen. Welche Welten das aber sind, ist bis hierher offen.

ähnlichster p-Welten. Vgl. Lewis (1973a: 19–21), List und Menzies (2009: 483), Kroedel (2019: 32–33).

[53] Lewis (1979: 465).

Was also fehlt, sind inhaltliche Maßstäbe für die Ähnlichkeit zwischen möglichen Welten. Lewis macht im obigen Zitat deutlich, dass hier keine allgemeingültigen Maßstäbe zu erwarten sind. Stattdessen werden die Maßstäbe für Ähnlichkeit von Kontext zu Kontext variieren. Dies sei im Lichte der Vagheit und Kontextabhängigkeit von kontrafaktischen Konditionalen zu erwarten: Die Ungenauigkeit in der Analyse spiegelt demnach eine Ungenauigkeit in der Sache wider.[54]

Zur Verdeutlichung vergleiche man die folgenden beiden kontrafaktischen Konditionale:[55]

> **(KK1)** Wenn Cäsar das Kommando im Koreakrieg gehabt hätte, dann hätte er mit Katapulten gekämpft.
>
> **(KK2)** Wenn Cäsar das Kommando im Koreakrieg gehabt hätte, dann hätte er mit Atomwaffen gekämpft.

Welches dieser kontrafaktischen Konditionale ist wahr? Dies hängt nach den eingeführten Wahrheitsbedingungen kontrafaktischer Konditionale davon ab, wie wir die Ähnlichkeit zwischen möglichen Welten bestimmen. Wenn Welten, in denen Cäsar als Kommandant im Koreakrieg mit Atomwaffen kämpft, der aktualen Welt ähnlicher sind als Welten, in denen Cäsar als Kommandant im Koreakrieg mit Katapulten kämpft, ist KK1 wahr und KK2 falsch. Wenn Welten, in denen Cäsar im Koreakrieg mit Katapulten kämpft, der aktualen Welt ähnlicher sind als Welten, in denen Cäsar im Koreakrieg mit Atomwaffen kämpft, ist KK2 wahr und KK1 falsch. Je nachdem, welchen Maßstab wir für die Ähnlichkeit zwischen Welten ansetzen, können jedoch beide Ähnlichkeitsurteile plausibel werden: Legen wir mehr Wert auf Ähnlichkeit mit Blick auf Cäsars aktualen Charakter als mit Blick auf Cäsars aktuale Waffenausstattung, scheinen Welten, in denen Cäsar mit Atomwaffen kämpft, ähnlicher zu sein. Legen wir hingegen mehr Wert auf Ähnlichkeit mit Blick auf Cäsars aktuale Waffenausstattung, scheinen Welten, in denen Cäsar mit Katapulten kämpft, ähnlicher zu sein.

Wie Lewis betont, sind solche Uneindeutigkeiten nicht spezifisch für Urteile über die Ähnlichkeit zwischen möglichen Welten, sondern betreffen im Allgemeinen generische Ähnlichkeitsurteile:

> All this is not special to the comparative similarity of worlds that appears in my analysis of counterfactuals. It is the same sort of vagueness that arises if I say that Seattle

[54] Vgl. auch Lewis (1973a: 91–95).

[55] Das Beispiel stammt von Quine (1960: 222).

> resembles San Francisco more closely than it resembles Los Angeles. Does it? That depends on whether we attach more importance to the surrounding landscape, the architecture, the dominant industries, the political temper, the state of the arts, the climate, the public transportation system, the form of the city government, or what. Possible worlds are bigger than cities (sometimes), and are capable of differing in a greater variety of respects. They are also capable of being more alike than any two actual cities. Still, any problem posed by my use of comparative similarity differ only in degree, not in kind, from problems about similarity that we would be stuck with no matter what we did about counterfactuals. Somehow, we do have a familiar notion of comparative overall similarity, even of comparative similarity of big, complicated, variegated things like whole people, whole cities, or even – I think – whole possible worlds.[56]

Für generische Ähnlichkeitsurteile sind also eine Reihe von unterschiedlichen und in verschiedene Richtungen zerrenden Ähnlichkeits*hinsichten* relevant und es stellt sich die Frage, wie diese Ähnlichkeitshinsichten gewichtet werden.[57]

Man könnte nun einerseits bei dieser Einsicht stehenbleiben und damit zufrieden sein, dass sich die Ungenauigkeit kontrafaktischer Konditionale in der Ungenauigkeit von generischen Ähnlichkeitsurteilen über mögliche Welten widerspiegelt.[58] Jedoch kann diese Strategie im Rahmen einer kontrafaktischen Analyse von Kausalität nicht ganz zufriedenstellen. Denn die kontrafaktische Analyse von Kausalität ist darauf angewiesen, dass bestimmte kontrafaktische Konditionale *falsch* sind, deren Wahrheitswert durch eine intuitive Bewertung generischer Ähnlichkeit zwischen möglichen Welten nicht eindeutig bestimmt wird. Insbesondere wäre es schlecht für die Plausibilität der kontrafaktischen Analyse der Kausalität, wenn *rückwärtsgerichtete kontrafaktische Konditionale* wahr sind. In diesem Falle müsste die kontrafaktische Analyse Wirkungen als Ursachen ihrer Ursachen klassifizieren und könnte nicht zwischen Epiphänomenen und Wirkungen unterscheiden.

Rückwärtsgerichtete kontrafaktische Konditionale sind Konditionale, die sich im Antezedens auf ein Ereignis beziehen, das nach dem im Konsequenz behandelten Ereignis geschieht. Ein Beispiel:

> **(KK3)** Wenn die Scheibe zu t2 nicht zerbrochen wäre, dann hätte Susi zu t1 ihren Stein nicht geworfen.

[56] Lewis (1973a: 92).

[57] Vgl. für die Unterscheidung zwischen generischer Ähnlichkeit und Ähnlichkeitshinsichten auch Cowling (2016: 2–4).

[58] Dies scheint Lewis (1973a) Herangehensweise zu sein.

Dieses Konditional wirkt auf den ersten Blick nicht unplausibel: Wenn die Scheibe nicht zerbrochen wäre, hätte dies wohl daran liegen müssen, dass Susi ihren Stein nicht geworfen hat. Welten, in denen die Scheibe nicht zerbricht und Susi ihren Stein nicht wirft sind der aktualen Welt ähnlicher als Welten, in denen die Scheibe nicht zerbricht, obwohl Susi ihren Stein wirft. Für die kontrafaktische Analyse ergibt sich hieraus das Problem, dass wir folgerichtig den Bruch der Fensterscheibe als Ursache für Susis Wurf klassifizieren müssen. Dies würde das tatsächliche Ursache-Wirkungs-Verhältnis jedoch umkehren.[59] Zudem ergibt sich das Problem, dass Epiphänomene als Wirkungen klassifiziert werden müssen: Wenn das Barometer keinen niedrigen Luftdruck angezeigt hätte, dann hätte das wohl daran liegen müsse, dass es keinen Luftdruckabfall gegeben hat. Wenn es jedoch keinen Luftdruckabfall gegeben hätte, wäre es nicht zum Sturm gekommen. Also gilt auch: Wenn das Barometer keinen niedrigen Luftdruck angezeigt hätte, wäre es nicht zum Sturm gekommen.[60] Also ist der Barometerstand eine Ursache für den Sturm.

Es ist im Kontext der kontrafaktischen Analyse der Kausalität also wichtig, die Wahrheit rückwärtsgerichteter kontrafaktischer Konditionale wie (KK3) auszuschießen. Hierfür muss man sich auf Ähnlichkeitsmaßstäbe berufen, die solche rückwärtsgerichteten Konditionale falsch machen. Lewis schlägt in diesem Kontext das folgende Rezept für die Bewertung der Ähnlichkeit zwischen möglichen Welten vor:

(1) It is of the first importance to avoid big, widespread, diverse violations of law.
(2) It is of the second importance to maximize the spatiotemporal region throughout which perfect match of particular fact prevails.
(3) It is of the third importance to avoid even small, localized, simple violations of law.

[59] Man könnte stattdessen natürlich eine Forderung der zeitlichen Abfolge in die Definition von Kausalität einbauen. Dies wird jedoch oft vermieden, um die prinzipielle Möglichkeit von rückwärtsgerichteter Kausalität nicht per definitionem auszuschließen. Zudem wäre mit dieser Modifikation noch nicht das Problem gelöst, das rückwärtsgerichtete kontrafaktische Konditionale im Zusammenhang mit der Unterscheidung zwischen Epiphänomenen und Wirkungen aufwerfen.

[60] Dieser Schluss setzt so natürlich die Transitivität kontrafaktischer Konditionale voraus und ist deshalb nicht gültig. Die tatsächliche Argumentation müsste vorsichtiger vorgehen, bliebe jedoch auf das rückwärtsgerichtete Konditional angewiesen.

(4) It is of little or no importance to secure approximate similarity of particular fact, even in matters that concern us greatly.[61]

Naturgesetze spielen demnach eine wichtige Rolle für die Bewertung der Ähnlichkeit von möglichen Welten. Welten, in denen radikale Verstöße gegen die aktualen Naturgesetze weit verbreitet sind, sind der aktualen Welt weitaus weniger ähnlich als Welten, in denen es nur wenige und weniger radikale Abweichungen von den aktualen Naturgesetzen gibt. Das ergibt sich aus Bedingung (1).

Jedoch kann es für die Bewertung der relevanten kontrafaktischen Konditionale nicht vermieden werden, Welten zu betrachten, in denen es zumindest kleine Abweichungen von den aktualen Naturgesetzen gibt. Zumindest unter der Annahme, dass die aktuale Welt deterministisch ist, wird nämlich jede mögliche Welt mit denselben Naturgesetzen wie die aktuale Welt entweder zu jedem oder zu keinem Zeitpunkt exakt mit der aktualen Welt übereinstimmen.[62] Nur bei kleinen Abweichungen von den aktualen Naturgesetzen kann es überhaupt Raumzeitregionen geben, die eine perfekte Entsprechung in der aktualen Welt haben. Bedingung (2) führt entsprechend in aller Regel dazu, dass wir Welten betrachten, in denen es kleine Abweichungen von den aktualen Naturgesetzen gibt. Lewis spricht solche kleinen Abweichungen von den aktualen Naturgesetzen auch als ‘kleine Wunder’ an.

Jede Welt, die nur in einem bestimmten Zeitraum mit der aktualen Welt übereinstimmt, wird daher Abweichungen von den aktualen Naturgesetzen enthalten. Und Welten, die erst kurz vor dem Antezedens-Ereignisses von der aktualen Welt abweichen, sind der aktualen Welt ähnlicher als Welten, die *schon immer* von der aktualen Welt abweichen. Für die Bewertung des Konditionals ‘Wenn Susi den Stein nicht geworfen hätte, wäre die Scheibe nicht zerbrochen’ werden wir also Welten betrachten, die möglichst lange exakt mit der aktualen Welt übereinstimmen, und in denen kurz vor Susis Wurf ein ‘kleines Wunder’ – eine Abweichung von den aktualen Naturgesetzen – dazu führt, dass Susi den Stein nicht wirft.

Wenden wir dieses Rezept auf (KK3) an: ‘Wenn die Fensterscheibe nicht zerbrochen wäre, dann hätte Susi zuvor ihren Stein nicht geworfen’. Es gibt nun für

[61] Lewis (1979: 472).

[62] Geht man davon aus, dass die aktuale Welt nicht deterministisch ist, ist es noch immer unplausibel, dass es für jedes kontrafaktisches Konditional ein entsprechendes indeterminiertes Ereignis gibt. Dass die Welt nicht allgemein deterministisch ist, heißt nicht, dass es z. B. eine mögliche Welt mit den Naturgesetzen der aktualen Welt gibt, die der aktualen Welt bis kurz vor Susis Wurf exakt gleicht und erst dann von ihr abweicht. Vgl. auch J. Bennett (2003: 203–205).

die der aktualen Welt ähnlichsten Welten, in der die Fensterscheibe zerbricht und Susi ihren Stein nicht wirft, eine der aktualen Welt ähnlichere Welt, in der die Fensterscheibe zerbricht, obgleich Susi ihren Stein wirft. Grund hierfür ist Lewis' Bedingung (2): Welten, die erst nach Susis Wurf von der aktualen Welt abweichen, entsprechen der aktualen Welt in einer größeren Raumzeitregion exakt als Welten, die schon vor Susis Wurf von der aktualen Welt abweichen. Relevant für die Bewertung von (KK3) sind demnach Welten, in denen es erst nach Susis Wurf zu einem kleinen Wunder kommt. Entsprechend ist (KK3) unter Voraussetzung dieses Ähnlichkeitsmaßstabes falsch.[63]

Lewis' Rezept für die Bewertung der Ähnlichkeit zwischen möglichen Welten ist nicht frei von Problemen.[64] Und nicht alle Vertreter*innen einer kontrafaktischen Analyse gehen von diesem Rezept aus.[65] Es bleibt jedoch festzuhalten, dass die kontrafaktische Theorie der Kausalität nur unter der Voraussetzung plausibel wird, dass rückwärtsgerichtete kontrafaktische Konditionale falsch sind. Dies ist eine bedeutende Einschränkung für Ähnlichkeitsmaßstäbe, die im Rahmen der kontrafaktischen Analyse angewendet werden können: Nur solche Ähnlichkeitsmaßstäbe sind zulässig, die rückwärtsgerichtete kontrafaktische Konditionale unter gewöhnlichen Umständen falsch machen.

4.3.3 Ähnlichkeit: Zentrierung und Ersetzung

Aber selbst unter Voraussetzung dieser Einschränkung bleiben wichtige Fragen zur Bewertung der Ähnlichkeit zwischen möglichen Welten offen. Ich gehe im Folgenden noch auf zwei weitere Komplikationen im Zusammenhang mit Ähnlichkeitsmaßstäben für die Bewertung kontrafaktischer Konditionale im Kontext der kontrafaktischen Theorie der Kausalität ein: Die erste Komplikation betrifft die Frage, ob es Ähnlichkeitshinsichten gibt, die bei der Bewertung der generischen Ähnlichkeit zwischen möglichen Welten irrelevant sind. Dies hat Konsequenzen für die Bewertung faktischer kontrafaktischer Konditionale, die die in Abschnitt 4.3.12 besprochene Proportionalitätsforderung in der kontrafaktischen Theorie der Kausalität betreffen. Die zweite Komplikation betrifft die Frage, ob das Antezedens-Ereignis in den ähnlichsten möglichen Welten, in denen

[63] Vgl. auch Hüttemann (2013a: 109–110).

[64] Vgl. z. B. Paul und Hall (2013: 44–47) für weitere Kritik.

[65] Paul und Hall (2013: 48) gehen beispielsweise von einem Vergleich der Ähnlichkeit von *Weltausschnitten* zum Antezedenszeitpunkt aus, statt von einem Vergleich der Ähnlichkeit zwischen möglichen Welten.

es nicht stattfindet, *ersetzt* oder *entfernt* werden sollte. Dies hat Konsequenzen für die Entwicklung von Argumenten für mentale kontrafaktische Konditionale (Abschnitte 4.3.6. bis 4.3.10), den Status der drei Prinzipien der Verursachung (4.3.11.) und die Anwendung der Proportionalitätsforderung (4.3.12. und 4.3.13.).

Zur ersten Komplikation: Maßstäbe generischer Ähnlichkeit zwischen möglichen Welten ergeben sich aus einer Gewichtung von spezifischeren Ähnlichkeitshinsichten.[66] Muss jede Ähnlichkeitshinsicht in eine Bewertung generischer Ähnlichkeit einfließen? Lewis schreibt hierzu Folgendes:

> But first, a word of warning! Do not assume that just any respect of similarity you can think of must enter into the balance of overall similarity with positive weight. The point is obvious for some respects of similarity, if such there be. It contributes nothing to the similarity of two gemstones that both be grue. (To be grue is to be green and first examined before 2000 A.D. or blue and not first examined before 2000 A.D.) But even some similarities in less gruesome respects may count for nothing. They may have zero weight, at least under some reasonable resolution of vagueness. To what extent are the philosophical writings of Wittgenstein similar, overall to those of Heidegger? I don't know. But here is one respect of comparison that does not enter at all, not even with negligible weight: the ratio of vowels to consonants.[67]

Es gibt demnach durchaus Ähnlichkeitshinsichten, die für generische Ähnlichkeit keine Rolle spielen. Für die Frage, ob die Schriften Wittgensteins insgesamt den Schriften Heideggers ähneln, ist das Verhältnis von Konsonanten zu Vokalen in ihren Schriften nicht einmal von geringer Relevanz. Zumindest gibt es generische Ähnlichkeitsvergleiche, die diese Ähnlichkeitshinsicht nicht berücksichtigen. Dies legt nahe, dass Entsprechendes auch für generische Ähnlichkeit zwischen möglichen Welten gilt.

Wenn bestimmte Ähnlichkeitshinsichten für die Bewertung von generischer Ähnlichkeit zwischen Welten im relevanten Sinne irrelevant sind, eröffnet dies die prinzipielle Möglichkeit, dass es Welten gibt, die der aktualen Welt im selben Maße ähneln wie sie sich selbst ähnelt.[68] Alle Welten, die von der aktualen Welt ausschließlich in irrelevanten Hinsichten abweichen, sind der aktualen Welt

[66] Vgl. Lewis (2015: 20): „I usually explain overall similarity as a resultant, determined by the balance of very many similarities and dissimilarities in many respects, of which some weigh more heavily than others".

[67] Lewis (1979: 465–466).

[68] Lewis (1979: 466) weist in Reaktion auf Bowie (1979) darauf hin, dass die Annahme der starken Zentrierung durchaus aufrecht erhalten werden *kann*, wenn es irrelevante Ähnlichkeitshinsichten gibt: „I reply that there may not be any worlds that differ from ours only in respects that do not count, even if there are some respects that do not count." Es ist jedoch schwierig zu sehen, warum es solche maximal ähnlichen Welten in Anbetracht von Lewis'

maximal ähnlich. In diesem Fall wäre die relevante Ähnlichkeitsrelation zwischen möglichen Welten nicht stark, sondern nur schwach zentriert.

Der Unterschied zwischen starker und schwacher Zentrierung lässt sich wie folgt spezifizieren:

Starke Zentrierung: Eine generische Ähnlichkeitsrelation ist stark zentriert um die aktuale Welt genau dann, wenn gilt: Jede Welt w, die nicht mit der aktualen Welt identisch ist, ist der aktualen Welt *weniger ähnlich als* die aktuale Welt.

Schwache Zentrierung: Eine generische Ähnlichkeitsrelation ist schwach zentriert um die aktuale Welt genau dann, wenn gilt: Keine Welt w, die nicht mit der aktualen Welt identisch ist, ist der aktualen Welt *ähnlicher als* die aktuale Welt.

Anders ausgedrückt: Schwache Zentrierung, nicht aber starke Zentrierung, lässt zu, dass es Welten gibt, die der aktualen Welt maximal ähnlich sind. In beiden Fällen ist die aktuale Welt eine der Welten, die der aktualen Welt maximal ähnlich sind. Aber nur im Falle starker Zentrierung ist die aktuale Welt die *einzige* Welt, die der aktualen Welt maximal ähnlich ist.

Ob wir eine starke oder eine schwache Zentrierung für die relevante Ähnlichkeitsrelation ansetzen, hat wichtige Konsequenzen in erster Linie für die Bewertung faktischer kontrafaktischer Konditionale – also kontrafaktischer Konditionale mit wahrem Antezedens und wahrem Konsequenz. Unter Voraussetzung einer stark zentrierten Ähnlichkeitsrelation sind solche Konditionale immer wahr. Unter Voraussetzung einer schwach zentrierten Ähnlichkeitsrelation können sie hingegen falsch sein.

Betrachten wir zur Verdeutlichung das folgende faktische kontrafaktische Konditional:

(KK4) Wenn Susi den Stein geworfen hätte, dann wäre die Scheibe zerbrochen.

Unter der (hier vorausgesetzten) Annahme, dass Susi tatsächlich den Stein wirft und die Scheibe tatsächlich zerbricht, ist (KK4) ein faktisches kontrafaktisches Konditional. Zur Bewertung von (KK4) müssen wir die der aktualen Welt ähnlichsten Welten betrachten, in denen Susi den Stein wirft. Unter Voraussetzung einer stark zentrierten Ähnlichkeitsrelation ist hierfür allein die aktuale Welt relevant: Es gibt keine anderen Welten, in denen Susi den Stein wirft, und die der aktualen Welt genauso ähnlich sind wie die aktuale Welt. Da Susi also in der

---

Prinzip der freien Rekombination für mögliche Welten (vgl. Lewis (1986a: 87–88)) *nicht* geben sollte.

aktualen Welt den Stein wirft und die Scheibe in der aktualen Welt zerbricht, ist (KK4) wahr.

Unter Voraussetzung einer schwach zentrierten Ähnlichkeitsrelation können für die Bewertung von (KK4) jedoch weitere Welten relevant werden. Wenn es andere Welten gibt, die von der aktualen Welt nur in irrelevanten Hinsichten abweichen und die der aktualen Welt daher maximal ähnlich sind, und in denen Susi ebenfalls den Stein wirft, dann muss die Scheibe in *jeder* dieser maximal ähnlichen Welten zerbrechen, damit (KK4) wahr ist. Dass die Scheibe in der aktualen Welt zerbricht, ist nicht mehr hinreichend für die Wahrheit von (KK4). (KK4) drückt dann mehr aus, als dass Susi tatsächlich den Stein wirft und die Scheibe tatsächlich zerbricht. Vielmehr drückt (KK4) dann aus, dass die Scheibe auch in einer Reihe von anderen, maximal ähnlichen Welten zerbricht, in denen Susi den Stein wirft. Welche Welten das sind, hängt davon ab, welche Ähnlichkeitshinsichten als irrelevant eingestuft werden.

Dass in manchen Kontexten überhaupt eine schwach zentrierte Ähnlichkeitsrelation mit irrelevanten Ähnlichkeitshinsichten angesetzt werden sollte, kann vorläufig wie folgt motiviert werden: Erstens scheinen faktische kontrafaktische Konditionale nicht immer wahr zu sein. So könnte man meinen, dass faktische kontrafaktische Konditionale, die komplett unzusammenhängende Ereignisse verbinden, falsch sind: das faktische kontrafaktische Konditional 'Wenn Barack Obama 2008 zum Präsidenten der Vereinigten Staaten gewählt worden wäre, dann würde die durchschnittliche Entfernung zwischen Mond und Erde 384.400 km betragen' klingt zumindest schräg.[69] Zweitens können faktische kontrafaktische Konditionale scheinbar signifikante Informationen enthalten, wenn wir (noch) nicht wissen, ob ihr Antezedens wahr ist: 'Wenn Bernard Sanders 2020 zum Präsidenten der Vereinigten Staaten gewählt würde, dann würde er eine allgemeine Krankenversicherung einführen' ist auch dann informativ, wenn sich später herausstellt, dass Antezedens und Konsequenz wahr sind. Dies spricht dafür, dass das Konditional mehr aussagt, als dass das Antezedens und das Konsequenz wahr sind. Stattdessen wird hier eine Relevanz des Antezedens für das Konsequenz mitausgesagt.[70] Drittens können faktische kontrafaktische Konditionale in bestimmten theoretischen Kontexten möglicherweise wertvolle Arbeit leisten: Peter Menzies beispielsweise beruft sich auf faktische kontrafaktische Konditionale, um im Rahmen einer kontrafaktischen Theorie der Kausalität intuitiv problematische Kausalaussagen der Art 'Die Geburt meines Großvaters hat

[69] Vgl. Penczek (1997), Pernu (2016: 1038).

[70] Vgl. Pernu (2016: 1038).

meinen Tod verursacht' zu vermeiden.[71] Und wie wir sehen werden, sollen faktische kontrafaktische Konditionale laut einigen Autor*innen auch eine Rolle bei der Argumentation für mentale Verursachung im nicht-reduktiven Physikalismus spielen.

Diese Überlegungen weisen die Annahme einer schwachen Zentrierung also zumindest als prima facie attraktive theoretische Option aus. Wir sollten sie daher im Hinterkopf behalten: Sie wird von einigen Autor*innen in Anspruch genommen, um die Idee der Proportionalität im Rahmen der einfachen kontrafaktischen Theorie der Kausalität einzufangen.[72] Ich argumentiere allerdings in Abschnitt 4.3.14. dafür, dass die Annahme schwacher Zentrierung im Rahmen der kontrafaktischen Theorie der Kausalität schlecht motiviert ist und die theoretische Arbeit, die ihr von einigen Autor*innen zugewiesen wird, besser durch komplexe *gewöhnliche* kontrafaktische Konditionale geleistet werden kann.

Kommen wir zur zweiten Komplikation hinsichtlich der Wahl eines geeigneten Ähnlichkeitsmaßstabes im Kontext der kontrafaktischen Theorie der Kausalität: Hierbei geht es um 'gewöhnliche' kontrafaktische Konditionale mit falschem Antezedens. Betrachten wir noch einmal das im Kontext der einfachen kontrafaktischen Theorie der Kausalität ausschlaggebende kontrafaktische Konditional:

(KK5) Wenn Susi den Stein nicht geworfen hätte, wäre die Scheibe nicht zerbrochen.

Welche Welten sind für die Bewertung von (KK5) relevant? Dies hängt davon ab, wie die der aktualen Welt ähnlichsten Welten beschaffen sind, in denen Susi den Stein *nicht* wirft. Das Problem ergibt sich nun aus der Frage, was in solchen Welten *anstelle von* Susis Steinwurf stattfindet.

Gehen wir bloß von unqualifizierter, intuitiver Ähnlichkeit zwischen möglichen Welten aus, sollten jene Welten der aktualen Welt am ähnlichsten sein, in denen anstelle von Susis aktualem Steinwurf ein Ereignis stattfindet, das dem aktualen Steinwurf möglichst ähnlich ist. Diese Herangehensweise führt im Kontext der einfachen kontrafaktischen Theorie der Kausalität jedoch zu Problemen. Wieder zitiere ich zur Verdeutlichung David Lewis:

> What is the closest way to actuality for C not to occur? It is for C to be replaced by a very similar event, one that is almost but not quite C, one that is just barely over the border between versions of C itself and its nearest alternatives. But if C is taken to be fairly fragile, then, if almost-C occurred instead of C, very likely the effects of almost-C would be almost the same as the effects of C. So our causal counterfactual will

[71] Menzies (2004).

[72] Vgl. List und Menzies (2009), Pernu (2016).

> not mean what we thought it meant, and it may well not have the truth value that we thought it had. When asked to suppose counterfactually that C does not occur, we do not really look for the very closest possible world where C's conditions of occurrence are not quite satisfied. Rather, we imagine that C is completely and cleanly excised from history, leaving behind no fragment or approximation of itself.[73]

Das Problem besteht darin, dass (KK5) plausiblerweise falsch ist, wenn wir davon ausgehen, dass in den ähnlichsten Welten, in denen Susis Steinwurf nicht stattfindet, ein sehr ähnliches Ereignis stattfindet. Denn dieses sehr ähnliche Ereignis würde ebenfalls dazu führen, dass die Scheibe zerbricht. Wenn Susi zum Beispiel in einigen ähnlichsten Welten, in denen sie den Stein nicht wirft, einen anderen, sehr ähnlichen Stein wirft oder den Stein auf eine ein klein wenig andere Art und Weise wirft, kommt es in diesen Welten wohl noch immer zum Bruch der Scheibe. Die Falschheit von (KK5) würde laut der einfachen kontrafaktischen Theorie der Kausalität implizieren, dass Susis Wurf den Bruch der Scheibe nicht verursacht. Bei einer solchen Bewertung von (KK5) führt die einfache kontrafaktische Theorie also sogar in paradigmatischen Fällen von Verursachung zum falschen Ergebnis.[74]

Aus diesem Grund sprechen viele Autor*innen ein Verbot gegen solche *Ersetzungs-Evaluationen* der relevanten kontrafaktischen Konditionale aus. Anstatt Welten zu betrachten, in denen ein Ereignis stattfindet, das Susis aktualem Wurf sehr ähnelt, sollte man zur Bewertung von (KK5) demzufolge Welten betrachten, in denen Susis Wurf „is completely and cleanly excised from history".[75] In den Bewertungswelten für (KK5) findet dann weder Susis Wurf noch ein anderes Ereignis, das Susis Wurf auch nur entfernt ähnelt, statt. Nennen wir eine solche Bewertung eine *Entfernungs-Evaluation.* Unter Voraussetzung einer Entfernungs-Evaluation ist (KK5) wahr: In den ähnlichsten Welten, in denen Susi nicht wirft, findet nichts dergleichen statt und die Scheibe bleibt ganz.

---

[73] Lewis (2000: 190).

[74] Auch andere kontrafaktische Theorien, denen zufolge kontrafaktische Abhängigkeit zwischen distinkten Ereignissen nicht notwendig sondern nur hinreichend für Kausalität ist, sind von einem verwandten Problem betroffen: Wenn (KK5) und ähnliche kontrafaktische Konditionale falsch sind, wird die Beobachtung, dass kontrafaktische Abhängigkeit zwischen distinkten Ereignissen hinreichend für Kausalität ist, beinahe inhaltsleer, da sie so gut wie nie erfüllt ist. Vgl. hierzu auch Harbecke (2014: 366).

[75] Neben Lewis (2000: 190) diskutieren u. a. auch J. Bennett (1987: 369), Bennett (2003: 482), Lewis (1986b: 211), Collins et. al. (2004: 21), Paul und Hall (2013: 51) sowie Harbecke (2014: 366) eine entsprechende Einschränkung.

Was als Ersetzungs-Evaluation von (KK5) zählt, hängt offenbar auch davon ab, wie Susis aktualer Wurf individuiert wird bzw. welches modale Profil Susis aktualer Wurf hat. Insbesondere können wir Susis Wurf einerseits als ein sehr *modal zerbrechliches Ereignis* betrachten: Schon kleinste Änderungen in der genauen Bewegung Susis und des Steins würden dazu führen, dass *dieses* Ereignis nicht stattfindet. Andererseits können wir Susis Wurf als ein recht *modal stabiles Ereignis* betrachten: Es wäre noch dasselbe Ereignis, selbst wenn Susi den Stein anders werfen würde oder einen anderen Stein werfen würde.[76] Auf den ersten Blick ist eine Ersetzungs-Evaluation von (KK5) besonders dann katastrophal für die einfache kontrafaktische Theorie der Kausalität, wenn wir den aktualen Wurf als modal zerbrechlich betrachten. Gehen wir hingegen von einem modal stabilen Steinwurf aus, ist zumindest weniger klar, ob in den ähnlichsten Welten, in denen Susi den Stein nicht wirft, die Scheibe zu Bruch geht. Diese Beobachtung legt nahe, dass für die Vermeidung des genannten Problems der einfachen kontrafaktischen Theorie der Kausalität weniger eine allgemeine Vermeidung von Ersetzungs-Evaluation gefordert ist, als vielmehr die Wahl eines angemessenen modalen Profils der Ursachen-Ereignisse.

Es ist nicht ganz klar, worauf genau eine Entfernungs-Evaluation hinausläuft. Denn eine Entfernung von Susis Wurf ohne Ersetzung durch ein ähnliches Ereignis verlangt noch immer eine Ersetzung durch irgendein *unähnliches* Ereignis. Wodurch Susis Wurf aber in den Bewertungswelten für (KK5) ersetzt werden sollte, bleibt bis hierher offen. Jens Harbecke macht den radikalen Vorschlag, dass die Raumzeitregion, in der Susis Wurf stattfindet, in den Bewertungswelten schlicht *leer* sein sollte bzw. durch Vakuum ersetzt werden sollte.[77] Dies scheint jedoch eine Reihe von sehr unplausiblen kontrafaktischen Konditionalen wahr zu machen. Deutlich wird dies vielleicht, wenn wir zu kontrafaktischen Konditionalen über neurologische Ereignisse kommen: Eine Ersetzung der entsprechenden Hirnregionen, in denen bestimmte neurologische Ereignisse auftreten, mit Vakuum ist sicherlich nicht gesund für die betroffenen Personen. Es wäre aber nicht plausibel, allgemein kontrafaktische Konditionale der Art ‘Wenn neurologisches Ereignis n nicht stattgefunden hätte, würde es der entsprechenden Person gesundheitlich nicht gut gehen’ zu akzeptieren. Zumindest für solche Konditionale müssten wir dann wohl einen anderen Ähnlichkeitsmaßstab ansetzen.

---

[76] Vgl. für die Unterscheidung zwischen zerbrechlichen und stabilen Ereignissen Lewis (1986c).

[77] Harbecke (2014: 367).

Ein besserer Vorschlag stammt von Ned Hall und Laurie Paul: Die Bewertung von (KK5) und anderen relevanten kontrafaktischen Konditionalen setzt eine Annahme darüber voraus, was *normalerweise* in der Raumzeitregion stattfinden würde, in der Susis Wurf stattfindet. Susis Wurf wird demnach als eine Abweichung von einem Normalverhalten aufgefasst. Ein im gegebenen Kontext naheliegendes Normalverhalten wäre dabei, dass Susi stillsteht, anstatt den Stein zu werfen. Zur Bewertung von (KK5) sollten wir also Welten betrachten, in denen Susi stillsteht.[78]

Die Unterscheidung zwischen Ersetzungs- und Entfernungs-Evaluationen kontrafaktischer Konditionale spielt eine wichtige Rolle für meine spätere Diskussion von mentaler Verursachung im Rahmen der kontrafaktischen Theorie der Kausalität. Beim jetzigen Stand bleibt zunächst festzuhalten, dass es erstens prima facie Gründe gibt, Ersetzungs-Evaluationen im Rahmen der kontrafaktischen Theorie der Kausalität zu vermeiden und dass zweitens Entfernungs-Evaluationen zunächst offenlassen, was in den Bewertungswelten die Stelle des entfernten Ereignisses einnimmt.

4.3.4 Probleme der einfachen kontrafaktischen Theorie

Die Schwierigkeit, einen geeigneten Ähnlichkeitsmaßstab für die Bewertung der für die einfache kontrafaktische Theorie der Kausalität relevanten kontrafaktischen Konditionale zu finden, stellt sicherlich ein großes Problem für diese Theorie dar. Es wird uns in den kommenden Abschnitten weiter beschäftigen. In diesem Abschnitt möchte ich jedoch zunächst knapp auf zwei verwandte und weit diskutierte Einwände gegen die einfache kontrafaktische Theorie der Kausalität eingehen, die weitgehend unabhängig von den diskutierten Problemen mit der Ähnlichkeitsrelation sind: Das erste Problem betrifft Fälle von Präemption und das zweite Problem betrifft Fälle von symmetrischer Überdetermination.

Zunächst zu Fällen von Präemption: Im bisher vorausgesetzten paradigmatischen Fall von Verursachung hängt der Bruch der Scheibe plausiblerweise von Susis Steinwurf kontrafaktisch ab. Die einfache kontrafaktische Theorie kommt daher in diesem Fall zu dem richtigen Ergebnis: Der Steinwurf verursacht den Bruch der Scheibe. In Fällen von Präemption kommt die einfache kontrafaktische Theorie jedoch zu einem falschen Ergebnis. Die folgende Abwandlung des paradigmatischen Falls kann dies veranschaulichen:

[78] Vgl. Paul und Hall (2013: 51 f.).

Frühe Präemption: Wieder wirft Susi ihren Stein und die Scheibe zerbricht. Neben Susi steht Paul, der ebenfalls einen Stein in der Hand hält. Wenn Susi nicht geworfen hätte, wäre Paul eingesprungen und hätte seinen Stein geworfen. In diesem Fall wäre die Scheibe noch immer zerbrochen.[79]

Bei früher Präemption ist das Urteil, dass Susis Wurf den Bruch der Scheibe verursacht, auf den ersten Blick nicht weniger plausibel als im paradigmatischen Fall von Verursachung. Jedoch hängt bei früher Präemption der Bruch der Scheibe nicht von Susis Wurf ab, weil es eine potentielle Ersatz-Ursache gibt: Pauls (möglicher) Wurf, der die Scheibe ebenfalls zerschmettert hätte.

Lewis weicht aufgrund von Fällen früher Präemption bereits in seiner ersten Formulierung einer kontrafaktischen Theorie der Kausalität von der einfachen kontrafaktischen Theorie der Kausalität ab.[80] Kontrafaktische Abhängigkeit zwischen distinkten Ereignissen sei nicht notwendig und hinreichend für Verursachung, sondern lediglich hinreichend. Frühe Präemption zeigt demzufolge, dass es Fälle von Verursachung gibt, in denen die Wirkung nicht kontrafaktisch von der Ursache abhängt. Kontrafaktische Abhängigkeit zwischen distinkten Ereignissen ist also nicht notwendig für Verursachung.

Stattdessen ist es nach Lewis früher kontrafaktischer Theorie notwendig für Verursachung, dass es eine Kette aus Ereignissen gibt, die Ursache und Wirkung verbinden und zwischen denen jeweils eine Beziehung der kontrafaktischen Abhängigkeit besteht. Dies ist im Fall früher Präemption gegeben: Der Bruch der Scheibe ist kontrafaktisch abhängig davon, dass Susis Stein sich auf halbem Weg zum Fenster befindet, und dies wiederum ist kontrafaktisch abhängig von Susis Wurf.

Jedoch lässt sich leicht eine Variation des Beispiels erdenken, in denen auch eine solche Kette aus kontrafaktisch abhängigen Ereignissen *nicht* vorliegt:

Späte Präemption: Susi und Paul werfen gemeinsam Steine auf Scheiben. Susi wirft nur ein klein wenig früher, so dass ihr Stein in der Scheibe einschlägt, kurz bevor Pauls Stein eingeschlagen wäre. Die Scheibe zerbricht.

Anders als im Fall früher Präemption macht Paul seinen Wurf in diesem Fall später Präemption nicht davon abhängig, ob Susi wirft: Er wirft nicht nur, wenn Susi nicht wirft, sondern auch, wenn Susi wirft. Tatsächlich wird der Prozess, der in Abwesenheit von Susis Wurf von Pauls Wurf zum Zerbrechen der Scheibe geführt hätte, erst kurz vor Vollendung unterbrochen: Erst kurz bevor Pauls Stein

[79] Vgl. Lewis (1973b), Hall (2004).

[80] Vgl. Lewis (1973b).

einschlagen würde, verhindert das Zerbrechen der Scheibe, dass Pauls Stein die Scheibe zerbricht.

Aus diesem Grund funktioniert Lewis Umgang mit früher Präemption in diesem Beispiel auch nicht: Das Zerbrechen der Scheibe ist nicht kontrafaktisch abhängig davon, dass sich Susis Stein auf dem halben Weg zur Scheibe befindet. Denn zeitgleich ist Pauls Stein (beinahe) auf dem halben Weg zur Scheibe. Und wenn Susis Stein nicht auf dem halben Weg zur Scheibe wäre, dann würde Pauls Stein noch immer die Scheibe zerschmettern.

Es gibt unterschiedliche Strategien, mit diesen und ähnlichen Gegenbeispielen gegen die einfache kontrafaktische Theorie umzugehen. Ich gehe nur ganz knapp hierauf ein:

Erstens kann man die Annahme, dass kontrafaktische Abhängigkeit *notwendig* für Kausalität ist, in einigen Zusammenhängen schlicht ohne Ersatz aufgeben. Gerade für die Diskussion der Frage nach der Existenz mentaler Verursachung im nicht-reduktiven Physikalismus ist die These, dass kontrafaktische Abhängigkeit *hinreichend* für Kausalität ist, ohnehin viel einschlägiger.[81] Diese These wird durch die Gegenbeispiele aber gar nicht angezweifelt.

Zweitens kann man die einfache kontrafaktische Theorie als gelungene Explikation *einer Art von Kausalität* verteidigen und zugleich darauf hinweisen, dass dies *nicht die einzige Art von Kausalität* ist.[82] Hier wird also die in Abschnitt 4.1. eingeführte Unterscheidung zwischen kausaler Abhängigkeit, hinreichender Verursachung und kausaler Produktion relevant. Diese Strategie geht mit der (zunächst kontraintuitiven) Annahme einher, dass Susis Wurf in den Präemptions-Fällen keine *Abhängigkeits*ursache für das Zerbrechen der Scheibe ist. Die widersprechenden Intuitionen können mit Verweis auf Produktions-Verursachung erklärt werden: Weil Susis Wurf eine paradigmatische Produktions-Ursache ist, sind wir geneigt, unreflektiert von einer ‚Ursache' zu sprechen. Führen wir uns jedoch die Unterscheidung zwischen Produktion und Abhängigkeit vor Augen und konzentrieren uns auf kausale Abhängigkeit, gewinnt das Urteil an Plausibilität: Schließlich macht Susis Wurf (in einem gewissen Sinne) tatsächlich keinen Unterschied dafür, ob die Scheibe zerbricht. Auch könnten wir das Zerbrechen der Scheibe nicht verhindern, indem wir Susi am Werfen hindern.

Drittens schließlich kann man die einfache kontrafaktische Theorie modifizieren. Eine Möglichkeit, dies zu tun, geht von der Beobachtung aus, dass die Abwesenheit von Susis Wurf in den Präemptions-Fällen dazu führt, dass die Scheibe nicht zerbricht, wenn bestimmte Faktoren in der Umgebung festgehalten

[81] Vgl. Kroedel (2019).

[82] Vgl. Hall (2004).

werden: So ist zwar das einfache kontrafaktische Konditional ‚Wenn Susi nicht geworfen hätte, dann wäre die Scheibe nicht zerbrochen' falsch. Das komplexere kontrafaktische Konditional ‚Wenn Susi nicht geworfen hätte *und Pauls Wurf (wie in der aktualen Situation) ebenfalls nicht getroffen hätte*, dann wäre die Scheibe nicht zerbrochen' hingegen ist wahr. Halten wir den aktualen Umstand, dass Pauls Wurf nicht trifft, also fest, ist der Bruch der Scheibe abhängig von Susis Wurf.[83] Notwendig für Kausalität ist demnach nicht simple kontrafaktische Abhängigkeit zwischen distinkten Ereignissen, sondern *de facto* kontrafaktische Abhängigkeit: kontrafaktische Abhängigkeit unter der Voraussetzung, dass bestimmte andere Faktoren festgehalten werden.

Verwandt mit Fällen von Präemption sind Fälle von symmetrischer Überdetermination. Diese waren im Zusammenhang mit dem Exklusionsprinzip bereits in Abschnitt 3.3.9. Thema. Dort ging es darum, dass die Möglichkeit symmetrischer Überdetermination eine Ausnahmeklausel im einfachen Exklusionsprinzip nötig macht. Fälle von symmetrischer Überdetermination stellen jedoch auch ein Problem für die einfache kontrafaktische Theorie der Kausalität dar.

> **Symmetrische Überdetermination**: Susi und Paul werfen gleichzeitig aus gleichem Abstand ihre Steine und die Steine treffen gleichzeitig auf die Scheibe. Die Scheibe zerbricht. Ein Einschlag hätte gereicht dafür, dass die Scheibe zerbricht.

Hier ist es – anders als im Fall später Präemption – plausibel, dass sowohl Susis Wurf als auch Pauls Wurf Ursachen für das Zerbrechen der Scheibe sind. Dennoch hängt das Zerbrechen der Scheibe von keinem der beiden Ursachen kontrafaktisch ab. Es handelt sich also um ein weiteres Gegenbeispiel gegen die These, dass kontrafaktische Abhängigkeit notwendig für Kausalität ist.

Die möglichen Reaktionen auf symmetrische Überdetermination entsprechen den möglichen Reaktionen auf Fälle von Präemption: Man kann erstens die These, dass kontrafaktische Abhängigkeit notwendig für Kausalität ist, ersatzlos streichen; man kann zweitens die einfache kontrafaktische Theorie der Kausalität auf kausale Abhängigkeit beschränken, akzeptieren, dass das Zerbrechen der Scheibe bei symmetrischer Überdetermination weder von Susis Wurf noch von Pauls Wurf abhängt und widersprechende Intuitionen durch Produktionsbeziehungen erklären; und man kann drittens die einfache kontrafaktische Theorie modifizieren.

[83] Vgl. Yablo (1992) und Paul und Hall (2013: 111–121) für weitere Diskussion.

Symmetrische Überdetermination ist für die Diskussion um mentale Verursachung im nicht-reduktiven Physikalismus deshalb besonders interessant, weil – zumindest im Kompatibilismus – auch mentale Verursachung im nicht-reduktiven Physikalismus in einem bestimmten Sinne mit symmetrischer Überdetermination einhergeht. Es ist jedoch wichtig zu sehen, dass sich die symmetrische Überdetermination, die durch den Fall von den gleichzeitigen Steinwürfen veranschaulicht ist, bedeutend unterscheidet von den Fällen von Überdetermination, die laut Kompatibilist*innen mit mentaler Verursachung einhergehen. Während der Punkt des Falls der gleichzeitigen Steinwürfe gerade ist, dass der Bruch der Scheibe von keinem der Steinwürfe kontrafaktisch abhängig ist, obgleich es sich um zwei Ursachen handelt, meinen Kompatibilist*innen, die sich auf die kontrafaktische Theorie der Kausalität berufen, dass mentale und physische Ereignisse einige Wirkungen überdeterminieren, weil die Wirkungen sowohl von mentalen Ereignissen als auch von physischen Ereignissen kontrafaktisch abhängen. Im einen Fall haben wir also Überdetermination *ohne kontrafaktische Abhängigkeit*, im anderen Fall Überdetermination *aufgrund von kontrafaktischer Abhängigkeit*. Dies ist ein bedeutender Unterschied.[84]

4.3.5 Direkte Argumente für die Existenz mentaler Verursachung im nicht-reduktiven Physikalismus

Um unter Voraussetzung der einfachen kontrafaktischen Analyse der Kausalität zu zeigen, dass es mentale Verursachung im nicht-reduktiven Physikalismus gibt, gilt es zu zeigen, dass entsprechende kontrafaktische Konditionale unter Voraussetzung des nicht-reduktiven Physikalismus wahr sind. Ich gehe in diesem Abschnitt zunächst von der (trivialen) Wahrheit *faktischer* kontrafaktischer Konditionale aus. Bezogen auf das Beispiel von Susi sind dann die folgenden gewöhnlichen kontrafaktischen Konditionale ausschlaggebend für die Begründung zugehöriger Kausalaussagen:

(KK6) Wenn Susi nicht die Idee für den Seminarbeitrag gehabt hätte, dann hätte sie nicht den Wunsch gehabt, den Arm zu heben.

(KK7) Wenn Susi nicht den Wunsch gehabt hätte, den Arm zu heben, dann hätte sie den Arm nicht gehoben.

---

[84] Vgl. auch Kroedel (2019: 170).

(KK8) Wenn Susi nicht den Wunsch gehabt hätte, den Arm zu heben, dann wären die mikrophysikalischen Ereignisse, die ihrer Armbewegung zugrunde liegen, nicht passiert.

Solche kontrafaktischen Konditionale, die sich an zentraler Stelle auf mentale Ereignisse beziehen, nenne ich im Folgenden *mentale kontrafaktische Konditionale*. Aus der Wahrheit von (KK6) können wir nach der kontrafaktischen Theorie der Kausalität schließen, dass Susis Idee ihren Wunsch verursacht, aus der Wahrheit von (KK7) können wir schließen, dass Susis Wunsch ihre Armbewegung verursacht und aus der Wahrheit von (KK8) können wir schließen, dass Susis Wunsch die mikrophysikalischen Ereignisse verursacht, die ihrer Armbewegung zugrunde liegen. Ein einfaches Argument für die Existenz mentaler Verursachung auf Grundlage der kontrafaktischen Theorie der Kausalität besteht demnach darin, für die Wahrheit geeigneter mentaler kontrafaktischer Konditionale zu argumentieren.

Das Schema eines solchen einfachen kontrafaktischen Arguments für mentale Verursachung im nicht-reduktiven Physikalismus kann wie folgt rekonstruiert werden:

Das einfache kontrafaktische Argument:
Es seien m ein mentales Ereignis und p* ein physisches Ereignis. m und p* treten tatsächlich ein und sind gänzlich distinkt. p* findet etwas später als m statt.

(EKA-1) Wenn m nicht eingetreten wäre, dann wäre p* nicht eingetreten.

(EKA-2) Wenn gilt: Wenn m nicht eingetreten wäre, dann wäre p* nicht eingetreten, dann gilt: m verursacht p*.

(EKA-K) Also: m verursacht p*.

(EKA-2) ist dabei eine Konsequenz aus der einfachen kontrafaktischen Theorie der Kausalität. Da es sich hier um die Behauptung handelt, dass kontrafaktische Abhängigkeit zwischen distinkten Ereignissen *hinreichend* für Kausalität ist, kann diese Prämisse nicht durch die in Abschnitt 4.3.3. angesprochenen Einwände gegen die einfache kontrafaktische Theorie angegriffen werden. Diese betreffen schließlich nur die Behauptung, dass kontrafaktische Abhängigkeit notwendig für Kausalität ist. Dennoch ist es natürlich möglich, dass andere Überlegungen auch gegen die Annahme sprechen, dass kontrafaktische Abhängigkeit zwischen distinkten Ereignissen hinreichend für Kausalität ist. Zunächst gehe ich aber davon aus, dass

Prämisse (EKA-2) wahr ist. Der Erfolg des einfachen kontrafaktischen Arguments ist dann allein von Prämisse (EKA-1) abhängig.

Was also spricht für die Wahrheit von (EKA-1)? Eine Möglichkeit, die Wahrheit der relevanten kontrafaktischen Konditionale zu etablieren, besteht darin, schlichtweg auf ihre hohe Ausgangsplausibilität zu verweisen.[85] Diese Strategie ist nach meiner Einschätzung stärker, als sie zunächst aussieht: Erstens kann man darauf verweisen, dass man sich bei der Bewertung kontrafaktischer Konditionale ohnehin ein Stück weit auf vortheoretische Plausibilitätserwägungen beziehen muss. Wie Lewis schreibt: "It is fair to discover the appropriate standards of similarity from the counterfactuals they make true, rather than vice versa"[86]. Um eine plausible Ähnlichkeitsrelation für die Bewertung kontrafaktischer Konditionale zu erhalten, müssen wir uns an vorgängigen Urteilen über die Wahrheit kontrafaktischer Konditionale orientieren. Typische Instanzen von (EKA-1) – wie 'Wenn Susi nicht den Wunsch gehabt hätte, den Arm zu heben, dann hätte sie den Arm nicht gehoben' – wirken plausibel genug, um ihre Wahrheit im Rahmen der Entwicklung einer Semantik kontrafaktischer Konditionale vorauszusetzen. Dies macht es von vornherein unwahrscheinlich, dass eine Begründung von (EKA-1) auf Grundlage einer Semantik kontrafaktischer Konditionale scheitert.

Zweitens gibt es keine spezifischen Argumente *gegen* die Wahrheit solcher kontrafaktischer Konditionale im nicht-reduktiven Physikalismus. Argumente, die allein auf Grundlage einer Semantik kontrafaktischer Konditionale und der Kernthesen des nicht-reduktiven Physikalismus zu etablieren suchen, dass mentale kontrafaktische Konditionale allgemein *falsch* sind, sind in der Literatur nicht zu finden. Es bleibt für Kritiker*innen des nicht-reduktiven Physikalismus allenfalls der Umweg über Exklusionsargumente: Da Exklusionsargumente etablieren, dass m keine Ursache für p* ist, kann – unter der Voraussetzung, dass kontrafaktische Abhängigkeit zwischen distinkten Ereignisse hinreichend für Kausalität ist – p* nicht kontrafaktisch abhängig von m sein. Zur Erwiderung können sich nicht-reduktive Physikalist*innen jedoch auf die in Kapitel 3 diskutierten Probleme von Exklusionsargumenten berufen: Die zentralen Exklusionsprinzipien sind letztlich schwer zu rechtfertigen. Zudem kann man darauf hinweisen, dass sich aus den Überlegungen zur kontrafaktischen Theorie der Kausalität eine weitere problematische Konsequenz von Exklusionsargumenten ergibt: Aus Exklusionsargumenten folgt – unter Voraussetzung einer kontrafaktischen Theorie der Kausalität – dass bestimmte

[85] Vgl. z. B. Baker (1993: 93).

[86] Lewis (1986b: 210–11).

kontrafaktische Konditionale falsch sind, die intuitiv wahr wirken und deren Bewertung als falsch nicht durch eine unabhängig plausible Semantik kontrafaktischer Konditionale gedeckt ist.

Der Verweis auf die Ausgangsplausibilität mentaler kontrafaktischer Konditionale kann also bereits einiges leisten. Dennoch ist es sinnvoll, zusätzliche *Argumente* für die Wahrheit der relevanten mentalen kontrafaktischen Konditionale zu betrachten. Dies hat die folgenden drei Gründe:

Erstens machen die zu besprechenden Argumente für die Wahrheit mentaler kontrafaktischer Konditionale im nicht-reduktiven Physikalismus deutlich, welche metaphysische Grundlage ihre Wahrheit hat: Sie zeigen genauer, aufgrund welcher fundamentalerer Umstände die mentalen kontrafaktischen Konditionale im nicht-reduktiven Physikalismus wahr sind.

Zweitens machen die Argumente deutlich, was mit einer Ablehnung der entsprechenden kontrafaktischen Konditionale verbunden wäre. Insbesondere wird in den zu besprechenden Argumenten die folgende Verbindung plausibilisiert: Wer die Wahrheit der relevanten mentalen kontrafaktischen Konditionale im nicht-reduktiven Physikalismus ablehnt, muss auch die Wahrheit einiger eng verbundener kontrafaktischer Konditionale über die Basen der im Antezedens erwähnten mentalen Ereignisse ablehnen. Diese Überlegung stärkt die Annahme, dass die relevanten mentalen kontrafaktischen Konditionale wahr sind.

Drittens können die zu behandelnden Argumente Kritiker*innen des nicht-reduktiven Physikalismus daher gewissermaßen abholen: Kritiker*innen des nicht-reduktiven Physikalismus wollen üblicherweise sagen, dass die entsprechenden kontrafaktischen Konditionale über physische Ereignisse wahr sind, insofern der nicht-reduktive Physikalismus falsch ist. Hierfür müssen sie sich aber auf eben jene Überlegungen stützen, die plausibilisieren, dass die entsprechenden kontrafaktischen Konditionale unter Voraussetzung des nicht-reduktiven Physikalismus wahr sind. Dann aber müssen sie akzeptieren, dass die physischen kontrafaktischen Konditionale auch im nicht-reduktiven Physikalismus wahr sind. Wenn hieraus aber folgt, dass auch die relevanten mentalen kontrafaktischen Konditionale wahr sind, müssen sie eben auch die Wahrheit der mentalen kontrafaktischen Konditionale im nicht-reduktiven Physikalismus akzeptieren.

Daher diskutiere ich in den nächsten Abschnitten Argumente für die Wahrheit von Prämisse (EKA-1). Die Grundidee der Argumente besteht darin, die Wahrheit mentaler kontrafaktischer Konditionale unter Bezug auf die Wahrheit von kontrafaktischen Konditionalen zu motivieren, die die *physischen Basen* der relevanten mentalen Ereignisse betreffen. Die im nicht-reduktiven Physikalismus angenommene enge modale Verknüpfung zwischen den Basen mentaler Ereignisse und den

zugehörigen mentalen Ereignissen erlaubt dabei, Verbindungen zwischen mentalen kontrafaktischen Konditionalen und kontrafaktischen Konditionalen über ihre Basen herzustellen. Ich diskutiere im kommenden Abschnitt ein Transitivitätsargument, das zwar nicht gültig ist, aber interessante Einsichten erlaubt. Hierauf aufbauend werden in den darauffolgenden Abschnitten zwei Ergänzungen des Transitivitätsarguments diskutiert.

4.3.6 Das ungültige Transitivitätsargument

Das Transitivitätsargument kann wir folgt formuliert werden:

Transitivitätsargument für mentale kontrafaktische Konditionale:
Es sei m ein mentales Ereignis. p sei die aktuale totale physische Basis von m. p* sei ein weiteres physisches Ereignis. m, p und p* treten tatsächlich ein. p* ist sowohl von m als auch von p gänzlich distinkt.

(TA-1) Wenn m nicht eingetreten wäre, dann wäre p nicht eingetreten.

(TA-2) Wenn p nicht eingetreten wäre, dann wäre p* nicht eingetreten.

(TA-K) Also: Wenn m nicht eingetreten wäre, dann wäre p* nicht eingetreten.

Die beiden Prämissen dieses Arguments sind aus Sicht des nicht-reduktiven Physikalismus sehr plausibel. (TA-1) folgt aus der Annahme, dass p – d. h. die physische Basis von m – m necessitiert: Es gilt notwendigerweise: Wenn p eintritt, dann tritt m ein.[87] Alle Welten, in denen m *nicht* eintritt, sind also auch Welten, in denen p *nicht* eintritt. Dies umfasst die nächsten Welten, in denen m nicht eintritt. Also sind die nächsten Welten, in denen m nicht eintritt, Welten, in denen auch p nicht eintritt. Also ist (TA-1) wahr. (TA-2) ist – zumindest für einige Einsetzungen von p* – ebenfalls überaus plausibel: Es gibt physische Ereignisse, die in ihrem Auftreten vom Auftreten der aktualen Basis von m abhängen.

Das Argument setzt jedoch voraus, dass kontrafaktische Abhängigkeit transitiv ist. Doch kontrafaktische Abhängigkeit ist nicht transitiv: Aus den kontrafaktischen

[87] Ich setze hier voraus, dass p als totale Basis aufgefasst werden kann. Die Argumente, die sich aus der Betrachtung totaler Basen ergeben, lassen sich plausiblerweise zu Argumenten ausbauen, die sich auf Kernbasen beziehen. Vgl. hierzu Zhong (2015: 165–170) und Kroedel (2019: 175–184).

Konditionalen ‘A > B’ und ‘B > C’ folgt nicht das kontrafaktische Konditional ‘A > C’. Der Grund hierfür ist, dass die nächsten A-Welten B-Welten sein können und die nächsten B-Welten C-Welten sein können, ohne dass die nächsten A-Welten C-Welten sind. Dies ist kann aber nur dann so sein, wenn die nächsten A-Welten nicht die *nächsten* B-Welten sind. Denn in diesem Fall können die nächsten B-Welten auch C-Welten sein, ohne dass die nächsten A-Welten C-Welten sind. Der transitive Schluss von ‘A > B’ und ‘B > C’ auf ‘A > C’ scheitert also typischerweise dann, wenn A eine relativ fernliegende Möglichkeit ist und B eine relativ naheliegende Möglichkeit ist. Zudem muss C in den relativ fernliegenden A-Welten falsch sein, aber in den näherliegenden B-Welten wahr sein.

Vielleicht kann ein Beispiel helfen, dies zu veranschaulichen. Lewis gibt das folgende Beispiel:

(KK9) Wenn Otto zur Party gegangen wäre, dann wäre Anna zur Party gegangen.

(KK10) Wenn Anna zur Party gegangen wäre, dann wäre Waldo zur Party gegangen.

(KK11) Wenn Otto zur Party gegangen wäre, dann wäre Waldo zur Party gegangen.

Wenn kontrafaktische Konditionale transitiv wären, könnten wir (KK11) aus (KK9) und (KK10) folgern. Jedoch lässt sich leicht einsehen, dass (KK11) falsch sein kann, obgleich (KK9) und (KK10) beide wahr sind: Otto und Anna sind enge Freunde, so dass Anna mit Otto zur Party gegangen wäre, wenn Otto nicht verhindert gewesen wäre. (KK9) ist daher wahr. Waldo mag Anna und hasst Otto. Er wäre deshalb zur Party gegangen, wenn Anna *ohne Otto* zur Party gegangen wäre. Otto konnte nicht zur Party, weil er sich zur Zeit auf der anderen Seite der Welt aufhält. Dass Otto zur Party geht, ist also eine fernliegende Möglichkeit. Anna hingegen wäre beinahe auch ohne Otto gegangen. Daher ist (KK10) wahr: In den nächsten möglichen Welten, in denen Anna zur Party geht, geht sie ohne Otto. In diesem Fall geht auch Waldo. Zugleich ist (KK11) falsch: Wenn Otto zur Party gegangen wäre, wäre Waldo ganz sicher nicht gegangen – selbst wenn Anna dann gegangen wäre.[88]

Das Beispiel funktioniert, weil Ottos Partybesuch eine relativ fernliegende Möglichkeit ist, während Annas Partybesuch eine näherliegende Möglichkeit ist: Die nächsten Welten, in denen Anna zur Party geht sind näher als die nächsten Welten, in denen Otto zur Party geht. Deswegen kann Waldo zwar in den nächsten Welten, in

[88] Vgl. Lewis (1973a: 32–33). Es gibt eine enge Verbindung zwischen Transitivtätsverletzung und Verletzungen von Monotonie. Das vorliegende Beispiel zeigt daher auch, dass aus ‚A > B‘ nicht folgt: ‚(A ∧ C) > B‘. Denn wenn Anna gegangen wäre, wäre Waldo gegangen. Aber es gilt nicht: Wenn Anna *und Otto* gegangen wären, wäre Waldo gegangen. Vgl. auch J. Bennett (2003: 160–161).

denen Anna zur Party geht, ebenfalls zur Party gehen, aber in den nächsten Welten, in denen Otto zur Party geht, fernbleiben, obwohl Anna in den nächsten Welten, in denen Otto zur Party geht, ebenfalls zur Party geht. Wenn Ottos Partybesuch hingegen eine ebenso naheliegende Möglichkeit ist wie Annas Partybesuch, funktioniert das Beispiel nicht mehr. Denn dann wären die nächsten Welten, in denen Otto zur Party geht, zugleich die nächsten Welten, in denen Anna zur Party geht. Wenn Waldo in diesen Welten zur Party geht, sind (KK9), (KK10) und (KK11) alle wahr.

Das Transitivitätsargument ist also deshalb nicht gültig, weil es sein kann, dass die nächsten Welten, in denen m nicht eintritt und die daher relevant für die Auswertung von (TA1) ('wenn m nicht eingetreten wäre, dann wäre p nicht eingetreten') sind, weiter entfernt sind als die nächsten Welten, in denen p nicht eintritt und die daher relevant für die Auswertung von (TA2) ('wenn p nicht eingetreten wäre, dann wäre p* nicht eingetreten') sind. Wenn in den Welten, die für (TA1) relevant sind, p* ohne p eintritt, während in den Welten, die für (TA2) relevant sind, p* nicht eintritt, sind (TA1) und (TA2) wahr, während (TAK) ('wenn m nicht eingetreten wäre, dann wäre p* nicht eingetreten') falsch ist.

Übertragen wir dies zur Veranschaulichung auf das Beispiel von Susi und betrachten dafür noch einmal das folgende Konditional:

> **(KK7)** Wenn Susi nicht den Wunsch gehabt hätte, den Arm zu heben, dann hätte sie den Arm nicht gehoben.

Das Transitivitätsargument würde (KK7) aus den folgenden beiden Konditionalen ableiten:

> **(KK12)** Wenn Susi nicht den Wunsch gehabt hätte, den Arm zu heben, dann wäre die aktuale Basis ihres Wunsches nicht eingetreten.

> **(KK13)** Wenn die aktuale Basis ihres Wunsches nicht eingetreten wäre, dann hätte sie den Arm nicht gehoben.

Dieser Schluss von (KK12) und (KK13) auf (KK7) funktioniert nicht, weil es sein könnte, dass die nächsten Welten, in denen Susis Wunsch nicht eintritt, weiter entfernt sind als die nächsten Welten, in denen die aktuale Basis von Susis Wunsch nicht eintritt. Dies mag – unter der Voraussetzung, dass Susis Wunsch multipel realisiert ist – recht plausibel sein: Die nächsten Welten, in denen die aktuale Basis von

Susis Wunsch nicht eintritt, sind dann Welten, in denen eine alternative Basis ihres Wunsches eintritt und ihr Wunsch somit noch immer eintritt.[89]

Es wird nun aber zugleich deutlich, welche modale Situation gegeben sein müsste, damit (KK7) falsch ist während (KK12) und (KK13) wahr sind: In den – relativ naheliegenden – nächsten Welten, in denen die aktuale Basis von Susis Wunsch nicht eintritt, hebt Susi nicht ihre Hand. In diesen Welten muss Susis Wunsch zugleich eintreten. Das Nicht-Auftreten von Susis Wunsch muss vielmehr eine fernliegendere Möglichkeit sein. Und in den[90] – also weiter entfernten – nächsten möglichen Welten, in denen Susis Wunsch *nicht* eintritt, müsste sie ihren Arm wieder heben. Dann wäre (KK7) falsch, während (KK12) und (KK13) wahr sind.

Diese modale Situation wirkt in Bezug auf die typischen Situationen, mit Bezug auf welche wir mentale kontrafaktische Konditionale äußern, nicht sehr plausibel: Warum sollten wir davon ausgehen, dass Susi zwar in den nächsten Welten, in denen die aktuale Basis ihres Wunsches nicht eintritt, während ihr Wunsch fortbesteht, ihren Arm *nicht* hebt, aber in den nächsten Welten, in denen sie ihren Wunsch *nicht* hat, den Arm wieder hebt?

Diese Seltsamkeit des Grundes für das Scheitern des Transitivitätsarguments mit Bezug auf die vorgestellte Situation legt nahe, dass man das Transitivitätsargument modifizieren oder um eine plausible Prämisse ergänzen kann, um zu einem gültigen und überzeugenden Argument zu gelangen. Und tatsächlich gibt es hier zumindest zwei Möglichkeiten, die in den folgenden beiden Abschnitten behandelt werden sollen.

4.3.7 Das Argument aus der wechselseitigen kontrafaktischen Abhängigkeit

Die erste Möglichkeit, das Transitivitätsargument zu modifizieren, besteht darin, es um die Prämisse zu ergänzen, dass die nächsten möglichen Welten, in denen die aktuale Basis von Susis Wunsch nicht eintritt, zugleich Welten sind, in der ihr Wunsch nicht eintritt. Anders ausgedrückt: Wenn die aktuale Basis von Susis Wunsch nicht eingetreten wäre, dann wäre auch Susis Wunsch nicht eingetreten. Es ergibt sich dann das folgende Argument:

[89] Dies wird im nächsten Abschnitt aber noch diskutiert.

[90] Genauer: ‚einigen der‘.

Argument aus der wechselseitigen kontrafaktischen Abhängigkeit:
Es sei m ein mentales Ereignis. p sei die aktuale totale physische Basis von m. p* sei ein weiteres physisches Ereignis. m, p und p* treten tatsächlich ein. p* ist sowohl von m als auch von p gänzlich distinkt.

(WA-1) Wenn m nicht eingetreten wäre, dann wäre p nicht eingetreten.

(WA-2) Wenn p nicht eingetreten wäre, dann wäre m nicht eingetreten.

(WA-3) Wenn p nicht eingetreten wäre, dann wäre p* nicht eingetreten.

(WA-K) Also: Wenn m nicht eingetreten wäre, dann wäre p* nicht eingetreten.

Dieses Argument ist – im Gegensatz zum Transitivitätsargument – gültig. Die zugrundeliegende Schlussregel besagt, dass aus 'A > B', 'B > A' und 'A > C' folgt: 'A > C'. Diese Schlussregel knüpft an die Einsicht an, dass A in Transitivitätsverletzungen eine relativ zu B fernliegendere Möglichkeit sein muss. Diese notwendige Bedingung für eine Transitivitätsverletzung wird durch die Annahme 'B > A' explizit ausgeschlossen.[91] Allgemein garantiert die Wahrheit von 'A > B' und 'B > A', dass die Menge der nächsten A-Welten identisch mit der Menge der nächsten B-Welten ist. Entsprechend sind die Welten, die relevant für die Auswertung von 'A > X' sind, identisch mit den Welten, die relevant für die Auswertung von 'B > X' sind. Deshalb sind A und B dann in Antezendenzien von kontrafaktischen Konditionalen unter Erhaltung des Wahrheitswertes austauschbar.[92]

Es stellt sich jedoch die Frage, ob wir die Gültigkeit des Arguments durch die Einführung einer inakzeptablen Prämisse erreicht haben. In 4.3.6. hatte ich bereits darauf hingewiesen, dass es aus Sicht des nicht-reduktiven Physikalismus zumindest prima facie plausibel ist, die nächsten Welten, in denen m nicht eintritt, als fernliegender zu betrachten als die nächsten Welten, in denen p, d. h. die aktuale physische Basis von m, nicht eintritt. Hintergrund ist die These der multiplen Realisierung: m ist in anderen möglichen Welten durch eine andere Basis als p realisiert. Insofern die nächsten Welten, in denen p nicht eintritt, solche Welten sind, in denen eine andere Basis von m eintritt, tritt m in diesen Welten noch immer ein. Die zentrale Prämisse (WA2) im Argument aus der wechselseitigen kontrafaktischen Abhängigkeit ('Wenn p nicht eingetreten wäre, dann wäre m nicht eingetreten') ist dann falsch.

---

[91] Vgl. Lewis (1973a: 33). Auch Kroedels (2015b) Argument für Abwärtsverursachung im Dualismus beruht auf dieser Schlussregel. Es unterscheidet sich vom vorliegenden Argument durch einen Bezug auf *die Menge aller möglichen Basen von m.* Das vorliegende Argument bezieht sich hingegen auf *die aktuale Basis von m.* Auf Kroedels Argumentationslinie gehe ich in Abschnitt 4.3.9. noch knapp ein.

[92] Vgl. J. Bennett (2003: 299–300).

An dieser Stelle wird jedoch erstmals eine der in Abschnitt 4.3.3. eingeführten Komplikationen relevant. Die Plausibilität von Prämisse (WA-2) hängt nämlich davon ab, ob wir eine Ersetzungsevaluation oder eine Entfernungsevaluation voraussetzen.

Unter Voraussetzung einer Ersetzungs-Evaluation ist (WA-2) plausiblerweise falsch: In den nächsten Welten, in denen die aktuale Basis von m nicht eintritt, tritt ein sehr ähnliches physisches Ereignis ein. Es ist plausibel, dass dieses sehr ähnliche Ereignis ebenfalls eine mögliche Basis von m ist. In diesen Welten tritt m also weiterhin ein. Es gilt also nicht: Wenn die aktuale Basis von m nicht eingetreten wäre, dann wäre m nicht eingetreten.

Man beachte jedoch, dass diese Überlegung verdächtig nahe an der in Abschnitt 4.3.3. geschilderten Überlegung ist, die dazu führt, dass die einfache kontrafaktische Theorie der Kausalität beinahe keine Ereignisse als Ursachen klassifiziert: In den nächsten Welten, in denen Susi den Stein nicht wirft, nimmt ein sehr ähnliches Ereignis den Platz von Susis Wurf ein. Dieses Ereignis könnte ein leicht anderer Wurf oder ein Wurf mit einem leicht anderen Stein sein. In diesen Welten würde die Scheibe noch immer zerbrechen. Das kontrafaktische Konditional 'Wenn Susi den Stein nicht geworfen hätte, wäre die Scheibe nicht zerbrochen' wäre dann falsch. Insofern diese Überlegung motiviert, Ersetzungsevaluationen allgemein zu vermeiden, wäre die eben geschilderte Argumentation für die Falschheit von (WA-2) also nicht überzeugend.[93]

Was ergibt sich für (WA2), wenn wir es stattdessen durch eine Entfernungsevaluation auswerten? Ich sehe keinen guten Grund, weshalb wir davon ausgehen sollten, dass p – die aktuale Basis von m – in den nächsten Welten, in denen es 'completely and cleanly excised from history'[94] ist, durch *eine andere Basis* von m ersetzt werden sollte. Folgen wir Harbeckes Ansatz, die Raumzeitregion von p in den Auswertungswelten durch Leere zu ersetzen, wird m in den Auswertungswelten sicherlich nicht mehr eintreten. Folgen wir dem Ansatz von Paul und Hall, p in den Auswertungswelten durch ein Normalverhalten zu ersetzen, ist es ebenfalls wenig naheliegend, das Eintreten einer alternativen Basis von m als ausschlaggebendes Normalverhalten zu betrachten. Wenn in den Auswertungswelten für (WA-2) also

[93] Man könnte hier jedoch einwenden, dass die geschilderte Überlegung mit Blick auf das Konditional ‚Wenn Susi den Stein nicht geworfen hätte, wäre die Scheibe nicht zerbrochen' lediglich ein Verbot von Ersetzungsevaluationen im Kontext von kontrafaktischen Konditionalen motiviert, die Informationen über Kausalbeziehungen ausdrücken sollen. Da es bei dem Konditional in (WA2) aber nicht unmittelbar um eine kausale Verknüpfung geht, könnten Ersetzungsevaluationen hier erlaubt sein.

[94] Lewis (2000: 190).

‘nothing remotely like’[95] p eintritt, wird dort m ebenfalls nicht eintreten.[96] Unter Voraussetzung einer Entfernungsevaluation ist (WA-2) also plausiblerweise wahr.

Da die beiden anderen Prämissen im Argument aus der wechselseitigen kontrafaktischen Abhängigkeit aus dem Transitivitätsargument übernommen wurden und dort bereits als überaus plausibel eingeschätzt wurden, schließe ich, dass das Argument aus der wechselseitigen kontrafaktischen Abhängigkeit überzeugen kann, insofern (WA-2) durch eine Entfernungsevaluation ausgewertet wird. Wir haben hier also bereits ein starkes Argument für das mentale kontrafaktische Konditional ‘Wenn m nicht eingetreten wäre, dann wäre p* nicht eingetreten’ und somit – unter der Annahme, dass dieses Konditional hinreichend für eine entsprechende Kausalitätsbehauptung ist – für die These, dass m p* verursacht.

4.3.8 Das Argument aus der eingeschränkten Transitivität

Es gibt aber noch eine weitere Möglichkeit, ausgehend von der Diagnose des Scheiterns des Transitivitätsarguments ein modifiziertes Argument für das entsprechende mentale kontrafaktische Konditional zu entwickeln. Selbst wenn wir nämlich zugestehen, dass die nächsten Welten, in denen m nicht eintritt, weiter entfernt sind als die nächsten Welten, in denen die aktuale Basis p nicht eintritt, müssen wir noch nicht zugestehen, dass p* in den nächsten Welten, in denen p nicht eintritt, ebenfalls nicht eintritt, aber in den nächsten Welten, in denen m nicht eintritt, wieder eintritt.

Da wir im nicht-reduktiven Physikalismus davon ausgehen können, dass p – als totale Basis von m – m necessitiert, gilt weiterhin, dass in jeder Welt, in der m *nicht* eintritt, auch p *nicht* eintritt. Wir gestehen nun zu, dass in den nächsten Welten, in denen p nicht eintritt, m noch immer eintritt. Auch gestehen wir zu, dass in diesen Welten p* nicht eintritt. Es lässt sich dann immer noch bestreiten, dass in den nächsten Welten, in denen m nicht eintritt (und p entsprechend weiterhin nicht eintritt) p* wieder eintritt. Und dies geht insbesondere auch, ohne die Konklusion, dass p* nicht eintreten würde, wenn m nicht eintreten würde, bereits vorauszusetzen. Allerdings müssen wir ein eng verwandtes Konditional voraussetzen – und dies kann die Stärke des angedachten Arguments in Zweifel ziehen.

---

[95] Wiederum Lewis (2000: 190).

[96] Das heißt allerdings nicht, dass alternative Basen von m ähnlich zu p sein müssen. Es ist lediglich nicht plausibel, dass p in den Auswertungswelten ausgerechnet durch alternative Basen von m ersetzt wird.

Das Argument kann wie folgt formuliert werden:

Argument aus der eingeschränkten Transitivität:
Es sei m ein mentales Ereignis. p sei die aktuale totale physische Basis von m. p* sei ein weiteres physisches Ereignis. m, p und p* treten tatsächlich ein. p* ist sowohl von m als auch von p gänzlich distinkt.

(ETA-1) Wenn m nicht eingetreten wäre, dann wäre p nicht eingetreten.

(ETA-2) Wenn m und p beide nicht eingetreten wären, dann wäre p* nicht eingetreten.

(ETA-K) Wenn m nicht eingetreten wäre, dann wäre p* nicht eingetreten.

Auch dieses Argument ist gültig. Die zugrundeliegende Schlussregel besagt, dass aus ‘$A > B$’ und ‘$(A \wedge B) > C$’ folgt: ‘$A > C$’.[97] Jonathan Bennett nennt diese Schlussregel eingeschränkte Transitivität (‘Limited Transitivity’).[98] Er argumentiert, dass diese Schlussregel auch erklären kann, wieso der transitive Schluss mit Blick auf kontrafaktische Konditionale in vielen Fällen recht plausibel wirkt. Beispielsweise wirkt dieser Schluss auf den ersten Blick einwandfrei:

(KK14) Wenn Susi den Stein geworfen hätte, wäre die Scheibe zerbrochen.

(KK15) Wenn die Scheibe zerbrochen wäre, hätte Susi Ärger bekommen.

(KK16) Also: Wenn Susi den Stein geworfen hätte, hätte sie Ärger bekommen.

Jedoch kann der Schluss von (KK14) und (KK15) auf (KK16) nicht durch eine Transitivitätsregel gedeckt sein. Stattdessen ist er plausiblerweise durch eingeschränkte Transitivität gedeckt. Denn (KK15) ist, so die Idee, im Kontext dieses Arguments eine Abkürzung für das komplexere Konditional ‘Wenn Susi den Stein geworfen hätte und die Scheibe (deshalb) zerbrochen wäre, hätte Susi Ärger bekommen’. (KK14) führt uns dann zu einer Möglichkeit, an die (KK15) anknüpft. Gerade dies ist in Transitivitätsverletzungen ja nicht der Fall: (KK11) (‘Wenn Otto zur Party gegangen wäre, wäre Susi gegangen’) führt uns zu einer Möglichkeit, die für die Auswertung von (KK12) (‘Wenn Susi zur Party gegangen wäre, wäre Waldo gegangen’) irrelevant ist. An (KK11) anknüpfen würde vielmehr das komplexere Konditional ‘Wenn Otto zur Party gegangen wäre und Susi (deshalb) zur Party gegangen wäre,

---

[97] Vgl. Lewis (1973a: 35).

[98] J. Bennett (2003: 298–301).

wäre Waldo zur Party gegangen'. Gerade dieses Konditional ist in dem Beispiel für Transitivitätsverletzung aber falsch.[99]

Was kann man also über die zentrale Prämisse (ETA-2) des Arguments aus der eingeschränkten Transitivität sagen? Was spricht aus Sicht des nicht-reduktiven Physikalismus dafür, dass gilt: Wenn m und p beide nicht eingetreten wären, wäre p* nicht eingetreten?

Man kann diese Prämisse durchaus durch die (unschuldigere) Prämisse (TA-2) aus dem Transitivitätsargument motivieren: Wenn p nicht eingetreten wäre, wäre p* nicht eingetreten. Zwar folgt (ETA-2) hieraus nicht, da kontrafaktische Konditionale nicht monoton sind: Aus '$A > B$' folgt nicht '$A \wedge C > B$'. Aber es ist eben auch nicht zu sehen, weshalb die Abwesenheit von m dazu führen sollte, dass p* wieder eintritt, obgleich p ja nach wie vor *nicht* eintritt.

Ein naheliegender Einwand gegen das Argument aus der eingeschränkten Transitivität besteht darin, dass die Prämisse (ETA-2) gefährlich nahe an der Konklusion ist. Es droht hier also Zirkularität. Man beachte jedoch, dass die Konklusion nicht aus (ETA-2) alleine folgt. Zudem kann (ETA-2) wie gerade gesehen ohne einen Verweis auf die Konklusion motiviert werden. Ich denke daher, dass das Argument aus der eingeschränkten Transitivität durchaus eine gewisse Überzeugungskraft entwickeln kann. Es verweist noch einmal auf die Seltsamkeit des Grundes für das Scheitern des Transitivitätsarguments. Im Gegensatz zum Argument aus der wechselseitigen kontrafaktischen Abhängigkeit ist dieses Argument dabei nicht auf Ersetzungs- oder Entfernungsevaluationen angewiesen.

4.3.9 Das Argument aus der notwendigen Basiertheit

Ein weiteres Argument für die Wahrheit eines mentalen kontrafaktischen Konditionals der Art 'Wenn m nicht eingetreten wäre, dann wäre p* nicht eingetreten' wird in einer Reihe von neueren Veröffentlichungen von Thomas Kroedel verteidigt.[100] Anstatt – wie die bisherigen Argumente – die Wahrheit mentaler kontrafaktischer Konditionale in der Wahrheit kontrafaktischer Konditionale über *die aktuale physische Basis* von m zu begründen, geht Kroedel von der Wahrheit kontrafaktischer Konditionale über die Menge *aller möglichen physischen Basen* von m aus. Sein Argument kann wie folgt formuliert werden:

[99] Vgl. J. Bennett (2003: 298–301).

[100] Vgl. Kroedel (2015a), Kroedel (2015b) und Kroedel (2019).

Argument aus der notwendigen Basiertheit:

Es sei m ein mentales Ereignis. p* sei ein physisches Ereignis. m und p* treten tatsächlich ein. p* ist sowohl von m als auch von p gänzlich distinkt.

(ANB-1) Notwendig: m tritt genau dann ein, wenn eine der möglichen physischen Basen von m eintritt.

(ANB-2) Wenn keine der physischen Basen von m eingetreten wäre, dann wäre p* nicht eingetreten.

(ANB-K) Also: Wenn m nicht eingetreten wäre, dann wäre p* nicht eingetreten.

Das Argument ist gültig und beruht auf einer Schlussregel, derzufolge aus ‚$\Box(A \leftrightarrow B)$' und ‚$A > C$' folgt: ‚$B > C$'. Wenn zwei Propositionen einander necessitieren, sind sie im Kontext kontrafaktischer Konditionale unter Erhaltung des Wahrheitswertes austauschbar.[101]

Prämisse (ANB-1) motiviert Kroedel durch die von nicht-reduktiven Physikalist*innen akzeptierte Supervenienz-These. Dabei beruft er sich auf eine verbreitete Version der starken lokalen Supervenienz-These: Notwendigerweise gilt: Wenn ein Objekt o eine mentale Eigenschaft M hat, dann hat o eine physische Eigenschaft P, so dass notwendigerweise gilt: Jedes Objekt o‘, das P hat, hat auch M. Aus dieser Version von starker Supervenienz folgt tatsächlich, dass notwendigerweise jedes Objekt, das eine mentale Eigenschaft hat, auch eine physische Eigenschaft hat, die sie necessitiert. Unter der vorausgesetzten Kim’schen Konzeption von Ereignissen (und der Annahme, dass konstitutives Objekt und konstitutives Zeitintervall gleichbleiben) folgt die entsprechende These für Ereignisse: Für jedes mentale Ereignis m gibt es ein physisches Ereignis p, so dass aus dem Auftreten von p mit Notwendigkeit das Auftreten von m folgt.

In jeder Welt, in der m eintritt, tritt also ein zeitgleiches physisches Ereignis ein, das als Basis von m fungiert. Aus der These der multiplen Realisierung folgt jedoch, dass m nicht in jeder möglichen Welt *dieselbe* physische Basis hat. Das Eintreten von m necessitiert also nicht das Eintreten *der aktualen Basis* p von m. Dennoch lässt sich argumentieren, dass das Eintreten von m das Eintreten *einer möglichen Basis von* m necessitiert. Im Hintergrund steht dabei die Annahme,

[101] Diese Schlussregel ist gültig, insofern kontrafaktische Konditionale nicht hyperintensional sind. Wenn es eine Unterscheidung zwischen wahren und falschen kontra*möglichen* Konditionalen gibt, ist diese Bedingung nicht gegeben.

dass m notwendigerweise physisch basiert ist: m kann nicht ohne eine physische Basis eintreten.[102]

[102] Diese Annahme kann jedoch in Konflikt mit bestimmten Ausdeutungen der *Kontingenz* des Physikalismus geraten: Wenn der Physikalismus kontingent ist, sollte es mögliche Welten geben, in denen der Dualismus oder der Idealismus wahr sind. In solchen Welten treten mentale Ereignisse ohne eine physische Basis ein. Insofern auch das mentale Ereignis m in solchen Welten eintritt, necessitiert m nicht, dass eine *physische* Basis von m eintritt. In Reaktion auf dieses Problem könnte Kroedel erstens die Kontingenz des Physikalismus bestreiten. Zweitens könnte er bestreiten, dass in den entsprechenden dualistischen Welten dieselben mentalen Ereignisse eintreten, die tatsächlich eintreten. Drittens könnte er das Argument auf die Menge aller Basen von m (und nicht auf die Menge aller *physischen* Basen von m) beziehen. Tatsächlich würde sich an der Auswertung des kontrafaktischen Konditionals (ANB-2) hierdurch nicht groß etwas ändern: Denn die Welten, in denen m in einer nicht-physischen Basis basiert ist, sind laut nicht-reduktiven Physikalist*innen in jedem Falle weit entfernt. Die nächsten möglichen Welten, in denen keine Basis von m eintritt, dürften daher die nächsten möglichen Welten sein, in denen keine *physische* Basis von m eintritt. Die weit entfernten dualistischen und idealistischen Welten sind hingegen irrelevant für die Auswertung von (ANB-2*) (‚Wenn keine Basis von m eintreten würde, würde p* nicht eintreten'). Ich gehe daher davon aus, dass diese Überlegung zur Kontingenz des Physikalismus keine bedeutenden Auswirkungen für das Argument hat.

Die Überlegung zeigt auch, dass die These der starken Supervenienz in der von Kroedel verwendeten Modaloperator-Version in Konflikt mit der Kontingenz des Physikalismus steht. Es scheint mir daher sinnvoll, den ersten Notwendigkeitsoperator in der These der starken Supervenienz nicht im Sinne von metaphysischer Notwendigkeit, sondern im Sinne einer schwächeren Modalität zu lesen. Wichtig für die physikalistische Grundthese ist ohnehin der zweite Modaloperator, der entsprechend weiterhin im Sinne von metaphysischer Notwendigkeit gelesen werden sollte.

Die von Kroedel verwendete Version der Supervenienz-These unterscheidet sich übrigens in einer Hinsicht von der in Abschnitt 2.5.2. eingeführten Version starker lokaler Supervenienz: Während die dort eingeführte Version direkt über mögliche Welten quantifizierte, haben wir hier eine Formulierung, die Gebrauch von Modaloperatoren macht. Es mag hilfreich sein, die beiden Versionen starker lokaler Supervenienz einmal direkt gegenüberzustellen:

(Sup-L-Stark) Mentale Eigenschaften supervenieren genau dann lokal stark auf physischen Eigenschaften, wenn gilt: Für alle möglichen Welten w und w‘ und jedes Objekt o in w und o‘ in w‘ gilt: Wenn o in w dieselben physischen Eigenschaften hat wie o‘ in w‘, dann haben o und o‘ dieselben mentalen Eigenschaften.

(Sup-L-Stark*) Mentale Eigenschaften supervenieren genau dann lokal stark auf physischen Eigenschaften, wenn gilt: Notwendigerweise gilt: Für jedes Objekt o und jede mentale Eigenschaft M: Wenn o M instantiiert, dann gibt es eine physische Eigenschaft P, so dass gilt: o instantiiert P und notwendigerweise: Für jedes Objekt o’: Wenn o’ P instantiiert, dann instantiiert o’ auch M.

$\Box\, \forall o\, \forall M \in Men\, (Mo \rightarrow \exists P \in Phy\, (Po \wedge \Box\, \forall o'\, (Po' \rightarrow Mo')))$.

Ein Unterschied zwischen (Sup-L-Stark) und (Sup-L-Stark*) besteht darin, dass aus (Sup-L-Stark) nicht folgt, dass jedes Objekt, das eine mentale Eigenschaft hat, auch eine

Prämisse (ANB-2) ist – ebenso wie schon (TA-2) im Transitivitätsargument – für zumindest manche Einsetzungen von p* überaus plausibel: Man kann davon ausgehen, dass sich die physische Welt irgendwie anders weiterentwickelt hätte, wenn weder die aktuale Basis noch irgendeine andere physische Basis von m eingetreten wären.

Mit dem Argument aus der notwendigen Basiertheit steht also ein weiteres starkes Argument für die Wahrheit mentaler kontrafaktischer Konditionale im nicht-reduktiven Physikalismus bereit.

4.3.10 Mentale kontrafaktische Konditionale

Die Diskussion in den letzten vier Abschnitten hat gezeigt, dass es gute Gründe gibt, die Wahrheit zumindest mancher mentaler kontrafaktischer Konditionale im nicht-reduktiven Physikalismus zu akzeptieren. Ausgangspunkt der Argumente waren dabei durchgehend kontrafaktische Konditionale, die auf die eine oder andere Weise einen Bezug auf die *physischen Basen* der relevanten mentalen Ereignisse enthalten. Die Grundidee war also, dass die Wahrheit mentaler kontrafaktischer Konditionale unter Annahme der modalen Verpflichtungen des nicht-reduktiven Physikalismus aus der Wahrheit physischer kontrafaktischer Konditionale folgt. Wer also bestreiten möchte, dass mentale kontrafaktische Konditionale im nicht-reduktiven Physikalismus wahr sind, ist auch auf die Zurückweisung der entsprechenden physischen kontrafaktischen Konditionale festgelegt.

---

physische Eigenschaft hat. (Sup-L-Stark) macht eine Aussage über Objekte, die physisch ununterscheidbar sind, d. h. all ihre physischen Eigenschaften teilen. Diese sollen auch mental ununterscheidbar sein. Hieraus folgt aber eben nicht, dass alle Objekte, die eine mentale Eigenschaft haben, überhaupt eine physische Eigenschaft haben. Mit (Sup-L-Stark) ist es vielmehr kompatibel, dass es mögliche Objekte gibt, die mentale aber keine physischen Eigenschaften haben. Da alle Objekte, die keine physischen Eigenschaften haben, physisch ununterscheidbar sind, müssten jedoch alle Objekte ohne physische Eigenschaften notwendigerweise dieselben mentalen Eigenschaften haben. (Sup-L-Stark*) hingegen impliziert, dass notwendigerweise jedes Objekt, das eine mentale Eigenschaft hat, auch eine physische Eigenschaft hat. Beide Versionen von starker Supervenienz rufen daher – auf unterschiedliche Weisen – Bedenken in Zusammenhang mit der These der Kontingenz des Physikalismus hervor: Wenn dualistische oder idealistische Welten möglich sind, gibt es sowohl mögliche Objekte mit mentalen Eigenschaften, die keine physischen Eigenschaften haben, als auch mögliche physisch ununterscheidbare Objekte, die sich mental unterscheiden.

Diese Einsicht ist ein Fortschritt gegenüber dem in Abschnitt 4.3.5. angeschnittenen bloßen Verweis auf die Plausibilität mentaler kontrafaktischer Konditionale. Denn erstens wird deutlich, dass die Wahrheit mentaler kontrafaktischer Konditionale nicht mehr verlangt als die Wahrheit physischer kontrafaktischer Konditionale und die Wahrheit der modalen Verpflichtungen des nicht-reduktiven Physikalismus. Was immer also die physischen kontrafaktischen Konditionale wahrmacht, ist auch für die Wahrheit der zugehörigen mentalen kontrafaktischen Konditionale verantwortlich. Zweitens verdeutlichen die Argumente, dass die Wahrheit mentaler kontrafaktischer Konditionale von Kritiker*innen des nicht-reduktiven Physikalismus nicht angezweifelt werden kann, ohne dass auch die Wahrheit der genannten physischen Konditionale angezweifelt wird. Drittens schließlich akzeptieren Kritiker*innen des nicht-reduktiven Physikalismus typischerweise die genannten physischen Konditionale. Dies macht es schwer zu sehen, wie sie die mentalen kontrafaktischen Konditionale anzweifeln könnten. Insgesamt sollte die Wahrheit zumindest mancher mentaler kontrafaktischer Konditionale nach meiner Einschätzung also keine strittige Annahme mehr sein.

Recht unbestimmt blieb jedoch in meiner bisherigen Diskussion der Argumente, genau *welche* mentalen kontrafaktischen Konditionale durch sie begründet werden können. Welche physischen Ereignisse können für p* in den obigen Argumenten eingesetzt werden? Für welche physischen Ereignisse ist es plausibel, dass sie kontrafaktisch von der aktualen Basis von m bzw. vom Eintreten einer möglichen physischen Basis von m abhängen? Ist es insbesondere plausibel, typische vermeintliche Wirkungen mentaler Ereignisse wie andere mentale Ereignisse (Susis Wunsch) oder behaviorale Ereignisse (Susis Armheben) für p* einzusetzen? Oder dürfen wir nur *physische*$_{eng}$ Ereignisse wie die mikrophysikalischen Vorgänge, die Susis Armbewegung zugrunde liegen, für p* einsetzen?

4.3.11 Kontrafaktische Konditionale und der Status der drei Prinzipien der Verursachung im Stufenmodell der Realität

Die Beantwortung dieser Fragen möchte ich in diesem Abschnitt mit einer Diskussion des Status der drei Prinzipien der Kausalität in der einfachen kontrafaktischen Theorie der Kausalität verbinden. Denn die Überlegungen, die für die Beantwortung der am Ende des letzten Abschnitts genannten Fragen relevant sind, sind zugleich für die Bewertung der Prinzipien der Verursachung im Stufenmodell der Realität relevant. Und aus dem Status der drei Prinzipien der

Verursachung ergeben sich Konsequenzen für die Beantwortung der genannten Fragen.

Betrachten wir zunächst das Prinzip der Abwärtsverursachung:

> **Prinzip der Abwärtsverursachung**: Wenn ein Ereignis u eine Ursache für ein Ereignis w ist und Ereignis w ontologisch abhängig von einem Ereignis b ist, dann ist u auch eine Ursache für Ereignis b.

Es ist hilfreich, sich dem Status dieses Prinzips über die Diskussion eines verwandten Prinzips anzunähern, das sich – anstatt auf kausale Aussagen und Aussagen über ontologische Abhängigkeit – auf kontrafaktische Konditionale und Necessitationsbehauptungen bezieht:

> **Prinzip der Abwärtsabhängigkeit**: Wenn gilt: ‚Wenn u nicht eingetreten wäre, dann wäre w nicht eingetreten' und gilt: ‚Notwendigerweise: Wenn b eintritt, dann tritt w ein', dann gilt: ‚Wenn u nicht eingetreten wäre, dann wäre b nicht eingetreten'.
>
> $((\neg Eu > \neg Ew) \wedge \Box(Eb \rightarrow Ew)) \rightarrow (\neg Eu > \neg Eb)$[103]

Dieses Prinzip ist wahr. Denn wenn b w necessitiert, dann ist jede Welt, in der w nicht stattfindet, auch eine Welt, in der b nicht stattfindet. Wenn die nächsten Welten, in denen u nicht stattfindet, also Welten sind, in denen w nicht stattfindet, dann sind diese Welten auch Welten, in denen b nicht stattfindet. Wenn ein Ereignis w also von einem anderen Ereignis u kontrafaktisch abhängt, dann hängt auch jede totale Basis von w kontrafaktisch von u ab.[104]

Bezogen auf das Susi-Beispiel heißt das: Wenn gilt: ‚Wenn Susi nicht den Wunsch gehabt hätte, den Arm zu heben, dann hätte sie ihren Arm nicht gehoben', und die mikrophysikalischen Vorgänge, die Susis Armbewegung zugrunde liegen, ihre Armbewegung necessitieren, dann gilt auch: ‚Wenn Susi nicht den Wunsch gehabt hätte, den Arm zu heben, dann hätten die mikrophysikalischen Vorgänge, die ihrer Armbewegung zugrunde liegen, nicht stattgefunden'.

Wenn wir von der einfachen kontrafaktischen Theorie der Kausalität ausgehen (und ontologische Abhängigkeit für den Moment mit Necessitation gleichsetzen), ist das Prinzip der Abwärtsabhängigkeit äquivalent zum Prinzip der Abwärtsverursachung. Die einfache kontrafaktische Theorie der Kausalität bestätigt also das Prinzip der Abwärtsverursachung.

---

[103] ‚Eu' steht hier für: ‚Ereignis u tritt ein'.

[104] Vgl. auch Zhong (2011: 141). Allgemein gilt: Aus ‘A > B’ und ‘□(B → C)’ folgt ‘A > C’.

Gehen wir hingegen aufgrund der in Abschnitt 4.3.4. geschilderten Probleme davon aus, dass kontrafaktische Abhängigkeit zwischen distinkten Ereignissen zwar hinreichend aber nicht notwendig für Kausalität ist, folgt das Prinzip der Abwärtsverursachung nicht mehr unmittelbar aus dem Prinzip der Abwärtsabhängigkeit. Denn es kann dann Kausalbeziehungen geben, bei denen die Wirkungen nicht von ihren Ursachen kontrafaktisch abhängen. Und für diese Kausalbeziehungen hat das Prinzip der Abwärtsabhängigkeit keine Relevanz. Dennoch können wir dann auf das Prinzip der Abwärtsabhängigkeit verweisen, um zumindest ein eingeschränktes Prinzip der Abwärtsverursachung zu rechtfertigen: Das Prinzip der Abwärtsverursachung gilt dann zumindest für solche Kausalbeziehungen, bei denen die Wirkungen von ihren Ursachen kontrafaktisch abhängen. In vielen typischen Fällen von Kausalität folgt also aus höherstufiger Verursachung Abwärtsverursachung: Wenn Susis Wunsch ihre Armbewegung verursacht, dann verursacht er auch die mikrophysikalischen Vorgänge, die ihrer Armbewegung zugrunde liegen.

Auch das Prinzip der Aufwärtsverursachung hat eine Entsprechung in einem verwandten Prinzip über kontrafaktische Konditionale und Necessitation:

Prinzip der Aufwärtsverursachung: Wenn ein Ereignis u eine Ursache für ein Ereignis w ist und ein Ereignis h ontologisch abhängig von Ereignis w ist, dann ist u auch eine Ursache für Ereignis h.

Prinzip der Aufwärtsabhängigkeit: Wenn gilt: ‚Wenn u nicht eingetreten wäre, dann wäre w nicht eingetreten' und gilt: ‚Notwendig: wenn w eintritt, dann tritt h ein', dann gilt: ‚Wenn u nicht eingetreten wäre, dann wäre h nicht eingetreten.

$((\neg Eu > \neg Ew) \wedge \Box(Ew \rightarrow Eh)) \rightarrow (\neg Eu > \neg Eh)$

Dieses Prinzip der Aufwärtsabhängigkeit ist jedoch nicht allgemein wahr.[105] Denn zwar ist – insofern w das höherstufige Ereignis h necessitiert – jede Welt, in der h nicht stattfindet, auch eine Welt, in der w nicht stattfindet. Jedoch kann es Welten geben, in denen w nicht stattfindet und h dennoch stattfindet. Wenn h multipel realisierbar ist, kann h auch durch ein anderes Ereignis als w realisiert sein. Und in Welten, in denen h durch ein anderes Ereignis realisiert ist, tritt h ohne w ein. Entsprechend können auch die nächsten Welten, in denen u nicht eintritt, solche Welten sein, in denen h anders realisiert ist. Dann wäre zwar das Konditional ‚Wenn u nicht eintreten würde, dann würde w nicht eintreten'

[105] Auch ist die zugehörige Schlussregel: ‚Aus ‚$A > B$' und '$\Box(C \rightarrow B)$' folgt: '$A > C$'' nicht gültig.

wahr, aber das Konditional ‚Wenn u nicht eintreten würde, dann würde h nicht eintreten' ist falsch.[106]

Ist es plausibel, dass in den für uns relevanten Situationen die nächsten Welten, in denen die Ursache nicht eintritt, Welten sind, in denen die Wirkung anders realisiert ist? Lei Zhong gibt das folgende Beispiel, um das Prinzip der Aufwärtsabhängigkeit zurückzuweisen:

> On the actual occasion, Ivan's pain is realized by P (the property of having such and such neurons firing in the anterior cortex), and Ivan's walking to the hospital is realized by P* (say, the property of having such and such fibers firing in the backbone). Suppose that Q, a physical property very similar to P, can also realize Ivan's pain, and Q*, a physical property very similar to P*, can also realize Ivan's walking. Since P* is caused by P, on a counterfactual theory of causation, in the closest possible worlds where P is not instantiated, P* is not instantiated either. But since it is reasonable to assume that Q, a realizer of pain, and Q*, a realizer of walking, would be also instantiated in this possible world, pain and walking would thus be instantiated in the same world.[107]

Zhongs Idee ist hier, dass in den nächsten Welten, in denen die aktuale Basis von Ivans Schmerz nicht eintritt, zwar die aktuale Basis von Ivans Gang zum Krankenhaus nicht eintritt, der Gang zum Krankenhaus jedoch noch immer eintritt, weil eine andere Basis des Gangs zum Krankenhaus eintritt. Zwar ist die aktuale Basis von Ivans Gang zum Krankenhaus von der aktualen Basis von Ivans Schmerz kontrafaktisch abhängig. Der Gang zum Krankenhaus selbst ist aber nicht von der aktualen Basis von Ivans Schmerz abhängig.

Wenn Zhong mit dieser Einschätzung recht hat, hat dies zwei für den vorliegenden Kontext relevante Konsequenzen: Erstens können in den in Abschnitten 4.3.5. bis 4.3.7. diskutierten Argumenten (Das Transitivitätsargument, das Argument aus der wechselseitigen Abhängigkeit und das Argument aus der eingeschränkten Transitivität) nur physische$_{eng}$ Basen höherstufiger Ereignisse für p* eingesetzt werden, nicht aber die höherstufigen Ereignisse selbst. Denn diese Argumente gingen von der Wahrheit physischer kontrafaktischer Konditionale aus, die die aktuale Basis eines mentalen Ereignisses betreffen. Das ausschlaggebende Konditional in diesen Argumenten ist ‘Wenn p – die aktuale Basis von m – nicht eingetreten wäre, dann wäre p* nicht eingetreten’. Nach Zhongs Einschätzung ist dieses Konditional falsch, wenn für p* ein typisches behaviorales Ereignis wie der Gang zu einem Krankenhaus eingesetzt wird. Wahr ist es hingegen, wenn für p* die aktuale physische Basis eines typischen behavioralen

[106] Vgl. Zhong (2011: 137–138).

[107] Zhong (2011: 137–138).

Ereignisses eingesetzt wird. Die Argumente in den Abschnitten 4.3.6. bis 4.3.8. könnten demnach bestenfalls etablieren, dass mentale Ereignisse physische$_{eng}$ Wirkungen haben. Sie könnten hingegen für sich genommen noch nicht etablieren, dass mentale Ereignisse diejenigen höherstufigen *behavioralen* Wirkungen haben, die wir ihnen typischerweise zuschreiben.

Zweitens können diese Argumente auch nicht durch einen Verweis auf das Prinzip der Aufwärtsabhängigkeit zu Argumenten für mental-behaviorale Verursachung ausgebaut werden. Denn dieses Prinzip ist laut Zhong nicht wahr. Daraus, dass die aktualen Basen behavioraler Ereignisse von mentalen Ereignissen kontrafaktisch abhängen, folgt nicht, dass die behavioralen Ereignisse selbst von mentalen Ereignissen kontrafaktisch abhängen. Die Wahrheit von kontrafaktischen Konditionalen der Art 'Wenn Susi den Wunsch, ihren Arm zu heben, nicht gehabt hätte, dann hätte sie ihren Arm nicht gehoben' muss dann auf andere Weise gerechtfertigt werden.

Kroedels Argument aus der notwendigen Basierheit scheint dieses Problem zu umgehen. Denn Zhong argumentiert nur dafür, dass höherstufige behaviorale Ereignisse nicht von *der aktualen Basis* mentaler Ereignisse kontrafaktisch abhängen. Ihr Eintreten kann aber noch immer von dem Eintreten *irgendeiner* physischen Basis des mentalen Ereignisses kontrafaktisch abhängen. Dies kann durchaus als großer Vorteil von Kroedels Argumentationslinie und ihrem Bezug auf die Menge der möglichen Basen gesehen werden.[108]

Ist Zhongs Argumentation gegen das Prinzip der Aufwärtsabhängigkeit aber plausibel? Dies hängt wiederum davon ab, ob wir eine Ersetzungs- oder eine Entfernungsevaluation der relevanten kontrafaktischen Konditionale ansetzen. Zhong verwendet für die physischen kontrafaktischen Konditionale, auf die es hier ankommt, offenbar eine Ersetzungsevaluation: Er geht davon aus, dass in den nächsten Welten in denen die aktuale Basis von Ivans Schmerz P nicht instantiiert ist, Q – „a property very similar to P“[109] – als alternative Basis von Ivans Schmerz instantiiert ist. Dies ist der Grund, warum er das kontrafaktische Konditional ‚Wenn die aktuale Basis von Ivans Schmerz nicht eingetreten wäre, dann wäre Ivan nicht ins Krankenhaus gegangen‘ für falsch hält: In den nächsten Welten, in denen die aktuale Basis des Schmerzes nicht eintritt, tritt eine sehr ähnliche Basis ein und Ivan geht noch immer ins Krankenhaus. Zugleich geht Zhong davon

---

[108] Kroedel meint jedoch, mit seinem Argument (nur?) Abwärtsverursachung zu rechtfertigen und bleibt hinsichtlich der Frage, welcherart Ereignisse in seinem Argument plausiblerweise für p* eingesetzt werden können, recht unbestimmt. Vgl. Kroedel (2015b: 844) und Kroedel (2019: 65–66) für knappe Bemerkungen.

[109] Zhong (2011: 137).

aus, dass das Konditional ‚Wenn die aktuale Basis von Ivans Schmerz nicht eingetreten wäre, dann wäre die aktuale Basis von Ivans Gang zum Krankenhaus nicht eingetreten' auch bei einer Ersetzungsevaluation wahr ist. Der Wechsel von P zu Q hätte dazu geführt, dass Ivans Gang zum Krankenhaus anders realisiert ist.[110]

Bei einer Entfernungsevaluation ist das Konditional ‚Wenn die aktuale Basis von Ivans Schmerz nicht eingetreten wäre, dann wäre Ivan nicht ins Krankenhaus gegangen' hingegen plausiblerweise wahr. Denn in den nächsten Welten, in denen die aktuale Basis von Ivans Schmerz ‚is completely and cleanly excised from history', tritt keine alternative Basis von Ivans Schmerz ein und Ivan geht nicht ins Krankenhaus. Allgemein ist zu erwarten, dass in Situationen, in denen kontrafaktische Konditionale der Art ‚Wenn p nicht eingetreten wäre, dann wäre p* nicht eingetreten' – wobei p und p* für $\text{physische}_{\text{eng}}$ Basen höherstufiger Eigenschaften stehen – auch die zugehörigen Konditionale wahr sind, die p mit den höherstufigen Eigenschaften verbinden, die von p* ontologisch abhängen, insofern alle Konditionale durch eine Entfernungsevaluation ausgewertet werden. Akzeptieren wir also ein Verbot von Ersetzungsevaluationen, scheint das Prinzip der Aufwärtsabhängigkeit durchaus plausibel.

Wieder folgt aus dem Prinzip der Aufwärtsabhängigkeit nur dann unmittelbar ein Prinzip der Aufwärtsverursachung, wenn wir voraussetzen, dass kontrafaktische Abhängigkeit zwischen distinkten Ereignissen hinreichend und notwendig für Kausalität ist. Wieder lässt sich aber auch ein eingeschränktes Prinzip der Aufwärtsverursachung aus dem Prinzip der Aufwärtsabhängigkeit ableiten, wenn wir die Annahme, dass kontrafaktische Abhängigkeit notwendig für Kausalität ist, aufgeben: In paradigmatischen Fällen von Verursachung wird daher – unter Voraussetzung von Entfernungsevaluationen – aus einer fundamentalen oder niedrigstufigen Kausalbeziehung eine aufwärtsgerichtete Kausalbeziehung folgen.

Kommen wir nun noch zum Prinzip der basalen Verursachung. Wieder diskutiere ich das Prinzip, in dem ich die Entsprechung betrachte, die sich auf kontrafaktische Konditionale und Necessitation bezieht:

> **Prinzip der basalen Verursachung**: Wenn ein Ereignis u eine Ursache für ein Ereignis w ist und Ereignis u ontologisch abhängig von einem Ereignis b ist, dann ist auch Ereignis b eine Ursache für Ereignis w.
>
> **Prinzip der basalen Abhängigkeit**: Wenn gilt: ‚Wenn u nicht eingetreten wäre, dann wäre w nicht eingetreten' und gilt: ‚Notwendig: Wenn b eintritt, dann tritt u ein', dann gilt auch: ‚Wenn b nicht eingetreten wäre, dann wäre u nicht eingetreten'.

[110] Zhong (2015: 161–163) akzeptiert diesen Einwand gegen seine Argumentation in Zhong (2011).

$((\neg Eu > \neg Ew) \wedge \Box(Eb \rightarrow Eu)) \rightarrow (\neg Eb > \neg Ew)$

Auch das Prinzip der basalen Verursachung ist nicht allgemein wahr.[111] Denn zwar sind, wenn b u necessitiert, alle Welten, in denen u nicht eintritt, zugleich Welten, in denen b nicht eintritt. Jedoch kann es wieder Welten geben, in denen u eintritt, aber b nicht eintritt. Auch hier ist die multiple Realisierbarkeit von u relevant: In einigen Welten ist u nicht durch b sondern durch eine andere physische Basis realisiert. In diesen Welten tritt b ohne u ein. Die nächsten Welten, in denen u nicht eintritt, sind also in jedem Falle Welten, in denen b nicht eintritt. ‚Wenn u nicht eingetreten wäre, dann wäre w nicht eingetreten' wird unter Betrachtung solcher Welten ausgewertet, in denen weder b noch u eintreten. Jedoch kann es nähere Welten geben, in denen b ebenfalls nicht eintritt, u aber eintritt. Und wenn es solche näheren Welten gibt, sind diese Welten relevant für die Auswertung von 'Wenn b nicht eingetreten wäre, dann wäre w nicht eingetreten'. Wenn w also in den nächsten Welten, in denen u (und b) nicht eintreten, ebenfalls nicht eintritt, aber in den nächsten Welten, in denen b nicht eintritt, w eintritt, ist das kontrafaktische Konditional 'Wenn b nicht eingetreten wäre, dann wäre w nicht eingetreten' falsch, während das kontrafaktische Konditional 'Wenn u nicht eingetreten wäre, dann wäre w nicht eingetreten' wahr ist, obgleich b u necessitiert.

Wieder spielt für die Begründung des Scheiterns des Schlusses von 'Wenn u nicht eingetreten wäre, dann wäre w nicht eingetreten' auf 'Wenn b nicht eingetreten wäre, dann wäre w nicht eingetreten', aber eine Ersetzungsevaluation eine ausschlaggebende Rolle. Ich zitiere wiederum Lei Zhong:

> It is reasonable to assume that any world where Ivan doesn't instantiate P and Ivan doesn't go to the hospital is more remote from the actual world than some world where Ivan instantiates Q rather than P (accordingly Ivan feels pain) and Ivan still goes to the hospital. So, P is not a good candidate for the cause of Ivan's going to the hospital. By contrast, a world where Ivan doesn't feel pain and doesn't go to the hospital is probably closer to the actual world than any world where Ivan doesn't feel pain but still somehow goes to the hospital.[112]

Die Zurückweisung des Prinzips der basalen Abhängigkeit beruht also auf ähnlichen Überlegungen wie die Zurückweisung des Prinzips der Aufwärtsabhängigkeit: Das Konditional 'Wenn die aktuale Basis von Ivans Schmerz nicht eingetreten wäre, dann wäre Ivan nicht ins Krankenhaus gegangen' ist falsch,

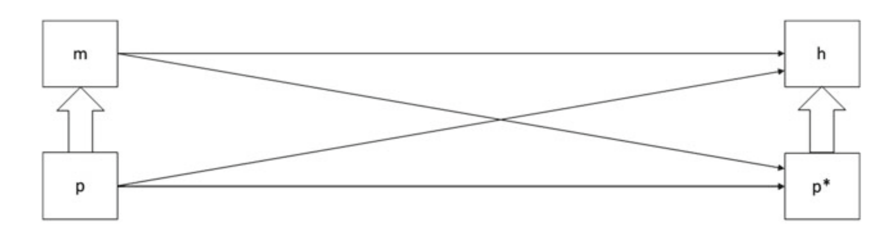
[111] Auch ist die zugehörige Schlussregel: ‚Aus ‚A > B' und '$\Box$(A → C)' folgt: 'C > B" nicht gültig.

[112] Zhong (2011: 134).

weil in den nächsten Welten, in denen die aktuale Basis nicht eintritt, die (sehr ähnliche) alternative Basis Q eintritt. Dies ist damit kompatibel, dass in den (weiter entfernten) nächsten möglichen Welten, in denen Ivans Schmerz nicht eintritt, Ivans Gang ins Krankenhaus ebenfalls nicht eintritt. Wie oben schon angemerkt würde eine Entfernungsevaluation zu einem anderen Ergebnis führen: Hier wäre das kontrafaktische Konditional 'Wenn die aktuale Basis von Ivans Schmerz nicht eingetreten wäre, dann wäre Ivan nicht ins Krankenhaus gegangen' plausiblerweise wahr, wenn das kontrafaktische Konditional 'Wenn Ivan keine Schmerzen gehabt hätte, wäre er nicht ins Krankenhaus gegangen' wahr ist. Auch das Prinzip der basalen Abhängigkeit ist also plausibel, wenn man ein Verbot von Ersetzungsevaluation akzeptiert. Entsprechendes gilt für die wie oben eingeschränkte Version des Prinzips der basalen Verursachung: In paradigmatischen Fällen von Kausalität, in denen die Wirkung kontrafaktisch abhängig von der Ursache ist, folgt daraus, dass ein höherstufiges Ereignis eine Ursache für ein Ereignis ist, dass auch die Basis des höherstufigen Ereignisses eine Ursache für dasselbe Ereignis ist.

Die Diskussion im diesem Abschnitt hat gezeigt, dass sich Kausalrelationen im nicht-reduktiven Physikalismus nach der kontrafaktischen Theorie der Kausalität stark vervielfältigen. Ausgehend von einer unschuldigen Kausalaussage wie 'Susis Wunsch verursacht ihre Armbewegung' können wir zunächst – da die kontrafaktische Theorie das Prinzip der Abwärtsverursachung bestätigt – schließen, dass ihr Wunsch auch alle totalen Basen der Armbewegung verursacht. Wenn wir zudem das für kontrafaktische Konditionale, die Informationen über Kausalverhältnisse ausdrücken sollen, plausible Verbot gegen Ersetzungsevaluationen akzeptieren, bestätigt die kontrafaktische Theorie zudem ein auf typische Fälle von Kausalität beschränktes Prinzip der Aufwärtsverursachung und ein auf typische Fälle von Kausalität beschränktes Prinzip der basalen Verursachung. Wenn Susis Wunsch also ein physisches$_{\text{eng}}$ Ereignis verursacht, dann verursacht er typischerweise auch alle höherstufigen Ereignisse, die durch dieses physische$_{\text{eng}}$ Ereignis necessitiert werden. Entsprechendes gilt für physische Ereignisse: wenn sie die Basis eines mentalen Ereignisses verursachen, verursachen sie auch das zugehörige mentale Ereignis. Und wenn Susis Wunsch ihre Armbewegung verursacht, wird typischerweise auch jede totale Basis von Susis Wunsch ihre Armbewegung verursachen. In typischen Situationen mentaler Verursachung ergibt sich daher das in Abbildung 4.1 dargestellte Bild.

Aus der kontrafaktischen Theorie der Kausalität folgt also soweit ein kompatibilistisches Bild mentaler Verursachung. Mentale Eigenschaften verursachen demnach (beinahe) ausschließlich Ereignisse, die auch physische$_{\text{eng}}$ Ursachen haben.



Erläuterung: m steht hier für ein mentales Ereignis wie Susis Wunsch, den Arm zu heben. h steht für eine höherstufige Wirkung des mentalen Ereignisse, wie das behaviorale Ereignis des Armhebens. p ist die physische$_{\text{eng}}$ Basis von m, p* die physische$_{\text{eng}}$ Basis von h. Die Pfeile stehen für die Relation der Verursachung im Sinne der einfachen kontrafaktischen Theorie ohne Proportionalitätsforderung.

Abbildung 4.1 Mentale Verursachung in der einfachen kontrafaktischen Theorie ohne Proportionalitätsforderung

Dies bringt das in 3.4.4. kurz angerissene Redundanzproblem auf den Plan. Zwar wurde nun gezeigt, dass mentale Ereignisse im nicht-reduktiven Physikalismus Ursachen sind. Allerdings sind sie nie *eigenständige* Ursachen, die sich von den physischen$_{\text{eng}}$ Ursachen, auf denen sie basieren, abheben. Um diesen Eindruck abzuschwächen, kann man sich nun auf eine Proportionalitätsforderung berufen. Die Einbindung einer Proportionalitätsforderung in die kontrafaktische Theorie der Kausalität ist daher Gegenstand des nächsten Abschnitts. Dabei orientiere ich mich an einer bestimmten Art, eine Proportionalitätsforderung in die kontrafaktische Theorie einzubinden, die von Christian List und Peter Menzies sowie Tuomas Pernu verteidigt wird.[113]

4.3.12 Proportionalität in der kontrafaktischen Theorie der Kausalität

Christian List und Peter Menzies setzen in ihrer Herangehensweise an mentale Verursachung im nicht-reduktiven Physikalismus eine einfache kontrafaktische Theorie der Kausalität voraus, der zufolge kontrafaktische Abhängigkeit zwischen

[113] List und Menzies (2009), Pernu (2016).

distinkten Ereignissen sowohl hinreichend als auch notwendig für Kausalität ist.[114] Den Kern dieser Theorie drücken sie wie folgt aus:

> *Truth conditions for making a difference*: The presence of F makes a difference to the presence of G in the actual world if and only if it is true in the actual world that (i) F is present □→ G is present; and (ii) F is absent □→ G is absent.[115]

Diese Formulierung ist recht nahe an der in Abschnitt 4.3.1. zu Beginn eingeführten Definition kontrafaktischer Kausalität (DKK). An die hiesige Terminologie angepasst und um eine Distinktheits- und Aktualitätsbedingung ergänzt,[116] können wir List und Menzies also die folgende Definition von Kausalität zuschreiben:

> **Definition kontrafaktischer Kausalität (DKK)**: Ein Ereignis u verursacht ein Ereignis w genau dann, wenn gilt: (dkk-i) u und w finden tatsächlich statt, (dkk-ii) u und w sind gänzlich distinkt und (dkk-iii) w ist kontrafaktisch abhängig von u, d.h. (k-i) wenn u eingetreten wäre, dann wäre w eingetreten und (k-ii) wenn u nicht eingetreten wäre, dann wäre w nicht eingetreten.

List und Menzies Idee ist nun, dass bereits diese einfache kontrafaktische Theorie der Kausalität implizit einer Proportionalitätsforderung Ausdruck verleiht, insofern die Ähnlichkeitsrelation, die für die Auswertung der relevanten kontrafaktischen Konditionale angesetzt wird, auf eine bestimmte Weise verstanden wird. List und Menzies Verständnis der Ähnlichkeitsrelation weicht dabei in einem wichtigen Punkt von der in den vergangenen Abschnitten vorausgesetzten Ähnlichkeitsrelation ab. Und zwar wird an dieser Stelle erstmals die in Abschnitt 4.3.3. geschilderte Komplikation mit Blick auf die Zentrierung der Ähnlichkeitsrelation relevant: Statt von einer starken Zentrierung der Ähnlichkeitsrelation auszugehen, setzen List und Menzies eine schwache Zentrierung voraus.

Dies hat wie geschildert Konsequenzen für die Auswertung faktischer kontrafaktischer Konditionale – also kontrafaktischer Konditionale mit wahrem Antezedens und wahrem Konsequenz. Während diese unter Voraussetzung starker Zentrierung trivial wahr sind, weil für ihre Auswertung ausschließlich die

[114] Die in Abschnitt 4.3.4. geschilderten Probleme der Präemption und Überdetermination übergehen die Autoren dabei, vgl. List und Menzies (2009: 484) (Fußnote 8) für eine entsprechende Anmerkung.

[115] List und Menzies (2009: 483). ‚□→‘ steht hier für das kontrafaktische Konditional. Ich verwende weiterhin das Zeichen '>' für das kontrafaktische Konditional.

[116] List und Menzies verweisen zwar nicht explizit auf diese Zusätze. Der Kontext in dem Aufsatz lässt es jedoch plausibel erscheinen, dass sie diese Zusätze akzeptieren würden.

aktuale Welt relevant ist, können sie bei schwacher Zentrierung falsch sein. Denn bei einer schwach zentrierten Ähnlichkeitsrelation kann es vorkommen, dass es maximal ähnliche Welten gibt, in denen das Antezedens eines faktischen kontrafaktischen Konditionals wahr ist, während das Konsequenz falsch ist. Die maximale Ähnlichkeit anderer Welten zur aktualen Welt ergibt sich dabei aus der Irrelevanz bestimmter Ähnlichkeitshinsichten. Maximal ähnliche nichtaktuale Welten sind dementsprechend Welten, die sich von der aktualen Welt ausschließlich in irrelevanten Hinsichten unterscheiden.

In der obigen Definition kontrafaktischer Kausalität kommt nun ein faktisches kontrafaktisches Konditional vor: Da u und w laut (dkk-i) tatsächlich eintreten, ist das kontrafaktische Konditional (k-i) ‚wenn u eingetreten wäre, dann wäre w eingetreten' ein faktisches kontrafaktisches Konditional. Unter Voraussetzung einer schwach zentrierten Ähnlichkeitsrelation ist (k-i) in der obigen Definition daher redundant. Die Wahrheit dieser Bedingung ist schon durch (dkk-i) garantiert. Dies erlaubt es, das faktische kontrafaktische Konditional für eine Auseinandersetzung mit aktualer Kausalität zu ignorieren. Und eben dies habe ich in den Abschnitten 4.3.4. bis 4.3.11. auch getan. Gehen wir jedoch – wie List und Menzies – von einer schwach zentrierten Ähnlichkeitsrelation aus, ist das Konditional ‚wenn u eingetreten wäre, wäre w eingetreten' auch dann nicht trivial wahr, wenn u und w tatsächlich eintreten. Es kann vielmehr Informationen ausdrücken, die für die Wahrheit (proportionaler) Kausalaussagen wesentlich sind.

Tatsächlich verstehen List und Menzies das faktische kontrafaktische Konditional in der Definition (DKK) so, dass es Ursachen ausschließt, die im Sinne einer Proportionalitätsforderung nicht spezifisch genug für ihre Wirkungen sind. Das gewöhnliche kontrafaktische Konditional soll hingegen Ursachen ausschließen, die zu spezifisch für ihre Wirkungen sind. Ich illustriere diese Idee zunächst unter Bezug auf das in 4.1.3. eingeführte Beispiel von Yablos Taube:

> **Yablos Taube**: Die Taube Sophie hat gelernt, auf alle roten und nur auf rote Objekte zu picken. Gerade legt jemand ein scharlachrotes Objekt vor Sophie und Sophie beginnt, zu picken.[117]

Hier ist der Umstand, dass ein *farbiges* Objekt vor Sophie gelegt wird, laut List und Menzies keine Ursache dafür, dass Sophie pickt, weil das faktische kontrafaktische Konditional ‚Wenn ein farbiges Objekt vor Sophie gelegt würde, würde

[117] Vgl. Yablo (1992: 257).

Sophie picken‘ falsch ist. Dieses Konditional schließt das Hinlegen eines farbigen Objekts also als proportionale Ursache für Sophies Picken aus, weil dieses Ereignis zu unspezifisch ist.

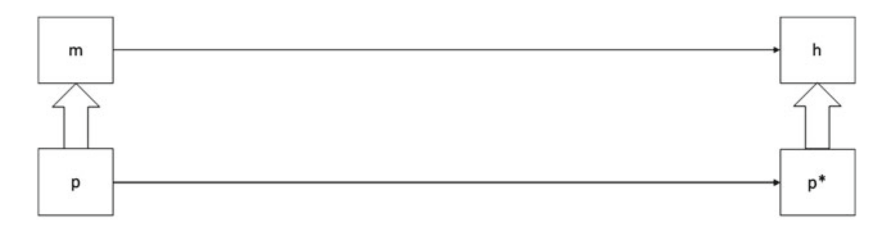
Das faktische kontrafaktische Konditional lädt laut List und Menzies dazu ein, nicht nur die aktuale Welt zu betrachten, sondern auch weitere Welten, in denen farbige Objekte vor Sophie gelegt werden. Dies sind die maximal ähnlichen Welten, die für die Auswertung des faktischen kontrafaktischen Konditionals relevant sind. Unter diesen maximal ähnlichen Welten sind auch solche, in denen ein farbiges, aber *nicht rotes* Objekt vor Sophie gelegt wird. In diesen Welten pickt Sophie jedoch nicht. Also ist das Konsequenz des faktischen kontrafaktischen Konditionals ‚Wenn ein farbiges Objekt vor Sophie gelegt würde, würde Sophie picken‘ nicht in allen maximal ähnlichen Welten wahr, in denen das Antezedens wahr ist. Also ist das faktische kontrafaktische Konditional falsch. Also ist das Legen eines farbigen Objekts vor Sophie keine Ursache für Sophies Picken.

Das Ereignis, das darin besteht, dass ein *scharlachrotes* Objekt vor Sophie gelegt wird, ist hingegen zu spezifisch, um als proportionale Ursache für Sophies Picken zu gelten. Hier argumentieren List und Menzies, dass das gewöhnliche kontrafaktische Konditional ‚Wenn kein scharlachrotes Objekt vor Sophie gelegt würde, würde Sophie nicht picken‘ falsch ist. Denn in einigen der nächsten möglichen Welten, in denen kein scharlachrotes Objekt vor Sophie gelegt wird, wird laut List und Menzies ein Objekt mit einem anderen Rotton vor Sophie gelegt. In solchen Welten, in denen zum Beispiel ein bordeauxrotes Objekt vor Sophie gelegt wird, pickt Sophie jedoch noch immer. Also ist das relevante kontrafaktische Konditional falsch. Die Annahme, dass in den nächsten möglichen Welten, in denen kein scharlachrotes Objekt vor Sophie gelegt wird, ein Objekt mit einem anderen Rotton vor Sophie gelegt wird, erinnert wiederum an die in Abschnitt 4.3.3. diskutierte Ersetzungsevaluation kontrafaktischer Konditionale.

Das Ereignis, das darin besteht, dass ein *rotes* Objekt vor Sophie gelegt wird, sei hingegen eine proportionale Ursache für Sophies Picken, weil sowohl das faktische kontrafaktische Konditional ‚Wenn ein rotes Objekt vor Susi gelegt würde, würde sie picken‘ als auch das gewöhnliche kontrafaktische Konditional ‚Wenn kein rotes Objekt vor Susi gelegt würde, würde sie nicht picken‘ wahr seien: In allen maximal ähnlichen Welten, in denen ein rotes Objekt vor Sophie gelegt wird, pickt sie und in den nächsten möglichen Welten, in denen kein rotes Objekt vor Sophie gelegt wird, pickt sie nicht.

4.3.13 Mentale Verursachung, Proportionalität und der Status der drei Prinzipien der Verursachung im Stufenmodell der Realität

Die so verstandene Proportionalitätsforderung kann nun auch auf typische Fälle mentaler Verursachung angewendet werden. Es stellt sich dabei heraus, dass alle drei Prinzipien der Verursachung im Stufenmodell der Realität falsch sind, wenn die Proportionalitätsforderung in der Ausdeutung von List und Menzies in Kraft ist. Denn die Proportionalitätsforderung führt zu einem Bild mentaler Verursachung, nach dem mentale Ereignisse ausschließlich höherstufige Ereignisse verursachen, ohne aber auch die Basen ihrer höherstufigen Wirkungen zu verursachen.[118] Im engen Sinne physische Ereignisse haben zudem zwar physische$_{\text{eng}}$ Wirkungen, verursachen aber eben nicht die höherstufigen Ereignisse, die von diesen physischen$_{\text{eng}}$ Wirkungen ontologisch abhängen. Wie in Abbildung 4.2 dargestellt ergibt sich demnach ein inkompatibilistisches[119] Bild mentaler Verursachung.



Erläuterung: m steht hier für ein mentales Ereignis wie Susis Wunsch, den Arm zu heben. h steht für eine höherstufige Wirkung des mentalen Ereignisse, wie das behaviorale Ereignis des Armhebens. p ist die physische$_{\text{eng}}$ Basis von m, p* die physische$_{\text{eng}}$ Basis von h. Die Pfeile stehen für die Relation der Verursachung im Sinne der einfachen kontrafaktischen Theorie mit Proportionalitätsforderung.

Abbildung 4.2 Mentale Verursachung in der einfachen kontrafaktischen Theorie mit Proportionalitätsforderung

Wie kann dieses Bild mentaler Verursachung ausgehend von der Proportionalitätsforderung begründet werden? Die Argumentation für das inkompatibilistische Bild mentaler Verursachung ergibt sich aus drei Argumenten: Einem Argument gegen Abwärtsverursachung, einem Argument gegen Aufwärtsverursachung und

[118] Vgl. Pernu (2016).
[119] Vgl. auch Abschnitt 3.4.1.

einem Argument für höherstufige Verursachung. Ich gehe die drei Argumente mit Bezug auf die Benennungen in Abbildung 4.2 durch.

Das Argument gegen Abwärtsverursachung kann – unter Bezug auf die Bezeichnungen in Abbildung 4.2 – wie folgt rekonstruiert werden:

Argument gegen Abwärtsverursachung:

(AB-1) Wenn m p* verursacht, dann gilt: Wenn m eintreten würde, dann würde p* eintreten.

(AB-2) Es gilt nicht: Wenn m eintreten würde, dann würde p* eintreten.

(AB-K) Also: m verursacht nicht p*.

(AB-1) besagt, dass das faktische kontrafaktische Konditional 'Wenn m eintreten würde, dann würde p* eintreten' notwendig für eine Kausalbeziehung ist. Diese Prämisse ergibt sich also unmittelbar aus List und Menzies Definition von Kausalität. Es ist zu beachten, dass die in Abschnitt 4.3.4. besprochenen Argumente gegen die Annahme, dass kontrafaktische Abhängigkeit notwendig für Kausalität ist, sich *nicht* gegen diese Prämisse (AB-1) richten. Denn diese Argumente betreffen das gewöhnliche kontrafaktische Konditional und nicht das hier genannte faktische kontrafaktische Konditional. Dass das faktische kontrafaktische Konditional notwendig für Kausalität ist, ist in der einfachen kontrafaktischen Theorie ohne Proportionalitätsforderung und mit starker Zentrierung eine triviale Konsequenz daraus, dass Ursache und Wirkung beide eintreten müssen. In der nun betrachteten kontrafaktischen Theorie mit Proportionalitätsforderung ist dies jedoch eine substantielle Bedingung an proportionale Ursachen.

Diese notwendige Bedingung ist laut List und Menzies (und Pernu) in ihrer Anwendung auf m und p* nicht erfüllt. Dies ist in Prämisse (AB-2) ausgedrückt. Der Grund dafür hat nun mit dem genaueren Verständnis der schwach zentrierten Ähnlichkeitsrelation zu tun. Denn die bloße Annahme, dass die Ähnlichkeitsrelation schwach zentriert ist, ist für sich genommen noch nicht sehr aussagekräftig. Wichtig ist, welche Ähnlichkeitshinsichten als irrelevant eingestuft werden und wie die Welten, die sich von der aktualen Welt nur in irrelevanten Hinsichten unterscheiden und daher maximal ähnlich sind, genauer beschaffen sind.

Die Ähnlichkeitshinsichten, die in der Bewertung des Konditionals 'Wenn m eingetreten wäre, dann wäre p* eingetreten' als irrelevant eingestuft werden sollten, sind laut List und Menzies Ähnlichkeiten mit Blick auf die physischen Basen von m. Welten, die sich nur darin von der aktualen Welt unterscheiden, dass in ihnen eine andere Basis von m eintritt, sollten für die Auswertung des Konditionals als

maximal ähnlich klassifiziert werden. Es reicht für die Wahrheit des Konditionals entsprechend nicht, dass m und p* in der aktualen Welt zusammen auftreten. Stattdessen müssen sie in allen Welten, die sich von der aktualen Welt nur dadurch unterscheiden, dass m eine andere physische Basis hat, zusammen auftreten. Die Beziehung zwischen m und p* muss in List und Menzies Worten 'realisierungsinsensitiv' sein: Änderungen an der Realisierung von m machen keinen Unterschied für p*.

Wenn wir jedoch davon ausgehen, dass p* ein physisches$_{eng}$ Ereignis ist, das als Basis für eine paradigmatische Wirkung von m (wie z. B. Susis Armbewegung) fungiert, ist es plausibel, dass die Beziehung zwischen m und p* *nicht* realisierungsinsensitiv ist. Dann wird p* in einigen der maximal ähnlichen Welten, in denen m eine andere Basis hat, nicht eintreten. Das Konditional 'Wenn m eintreten würde, dann würde p* eintreten' ist dann falsch und Prämisse (AB-2) im Argument gegen Abwärtsverursachung wahr.[120]

Es ist dabei anzumerken, dass dieses Argument nur in Fällen gegen Abwärtsverursachung spricht, in denen die Beziehung zwischen der vermeintlichen Ursache und der niedrigerstufigen vermeintlichen Wirkung nicht realisierungsinsensitiv ist. Es wird durch diese Argumentation also keinesfalls etabliert, dass mentale Ereignisse niemals physische$_{eng}$ Wirkungen haben. Wenn eine Person beispielsweise willentlich per Knopfdruck dafür sorgen kann, dass in einem Teilchenbeschleuniger zwei Elementarteilchen kollabieren, kann sein Wunsch, den Knopf zu drücken, durchaus eine proportionale Ursache für die entsprechenden mikrophysikalischen Ereignisse sein. Die Beziehung zwischen ihrem Wunsch, den Knopf zu drücken, und den mikrophysikalischen Ereignissen ist dann realisierungsinsensitiv: Es kommt zur Kollision der Elementarteilchen ganz gleich wie der Wunsch realisiert ist.

Das Argument gegen Aufwärtsverursachung kann – unter Bezug auf die in Abbildung 4.2 dargestellte Situation – wie folgt rekonstruiert werden:

Argument gegen Aufwärtsverursachung:

(AUF-1) Wenn p h verursacht, dann gilt: Wenn p nicht eintreten würde, dann würde h nicht eintreten.

(AUF-2) Es gilt nicht: Wenn p nicht eintreten würde, dann würde h nicht eintreten.

(AUF-K) Also: p verursacht nicht h.

Hier drückt Prämisse (AUF-1) aus, dass das gewöhnliche kontrafaktische Konditional 'Wenn p nicht eintreten würde, dann würde h nicht eintreten' notwendig für

[120] Vgl. Pernu (2016: 1040–1041).

Kausalität ist. Diese Annahme ist aufgrund der in Abschnitt 4.3.4. angesprochenen Beispiele von Präemption und Überdetermination eigentlich problematisch.[121] Man könnte in Reaktion hierauf vielleicht versuchen, sie auf geeignete Weise einzuschränken. Dies hatte ich in Abschnitt 4.3.11. bereits angedeutet: In ‘typischen’ Fällen, in denen *keine* genuinen Überdeterminierer oder Backup-Ursachen zu vermuten sind, erlaubt die Abwesenheit kontrafaktischer Abhängigkeit den Schluss auf die Abwesenheit von Kausalität. Da wir es im vorliegenden Fall nicht mit einem Fall von genuiner Überdetermination oder Präemption zu tun haben, können wir (AUF-1) akzeptieren.[122]

Prämisse (AUF-2) besagt, dass das relevante kontrafaktische Konditional falsch ist: In zumindest manchen der nächsten möglichen Welten, in denen p nicht eintritt, tritt h dennoch ein. Zur Begründung berufen sich sowohl List und Menzies als auch Pernu auf die schon aus Abschnitt 4.3.11. bekannte Ersetzungsevaluation des Konditionals: In den nächsten möglichen Welten, in denen p nicht eintritt, tritt eine alternative Basis von m ein. In diesen Welten tritt h noch immer ein. Wie schon in Abschnitt 4.3.11. angemerkt, ist diese Begründung auf die Ersetzungsevaluation angewiesen und funktioniert bei einer Entfernungsevaluation nicht.

Bis hierher ist die inkompatibilistische Argumentation rein negativ: Es liegen zwei Argumente *dagegen* vor, dass bestimmte Kausalbeziehungen bestehen. Der positive Teil der inkompatibilistischen Argumentation soll die Behauptung etablieren, dass m h verursacht. Das Argument für diese Behauptung kann wie folgt rekonstruiert werden:

Argument für höherstufige Verursachung:

> **(HV-1)** Wenn gilt: Wenn m eintreten würde, dann würde h eintreten und gilt: Wenn m nicht eintreten würde, dann würde h nicht eintreten, dann gilt: m verursacht h.
>
> **(HV-2)** Es gilt: Wenn m eintreten würde, dann würde h eintreten.
>
> **(HV-3)** Es gilt: Wenn m nicht eintreten würde, dann würde h nicht eintreten.
>
> **(HV-K)** Also: m verursacht h.

---

[121] Vgl. auch Moore (2017: 1470–1472) für eine Diskussion des Präemptionsproblems für die kontrafaktische Theorie mit Proportionalitätsforderung.

[122] Dies scheint der Ansatz zu sein, den auch List und Menzies (2009: 478) Fußnote 8 andeuten: „[W]e can set aside causal systems involving pre-emption and overdetermination as these complications are not germane to our discussion. Of course, if our examples could be understood as involving overdetermination, this simplifying assumption would be unwarranted“.

(HV-1) besagt, dass kontrafaktische Abhängigkeit zwischen distinkten Ereignissen – die ihren Ausdruck unter Voraussetzung der schwach zentrierten Ähnlichkeitsrelation sowohl im faktischen als auch im gewöhnlichen kontrafaktischen Konditional findet – *hinreichend* für Kausalität ist.

Über die Wahrheit des faktischen kontrafaktischen Konditionals in (HV-2) schreibt Pernu folgendes:

> [G]iven that m occurs some of its realisers must also occur, which in turn are causally sufficient for the realisers of h to occur, each of which in turn necessitate the occurrence of h.[123]

Die Idee ist hier also, dass die möglichen Basen von m Ursachen für die möglichen Basen von h sind und deshalb in allen Welten, in denen m andere Basen hat, eine Basis von h verursacht wird. Da die maximal ähnlichen Welten die Welten sind, in denen m andere Basen hat, und in jeder dieser Welten h eintritt, ist das Konditional 'Wenn m eingetreten wäre, dann wäre h eingetreten' wahr.

Auf eine ähnliche Weise argumentiert Pernu für die Wahrheit des gewöhnlichen kontrafaktischen Konditionals:

> [G]iven that m fails to occur, each of its realisers must also fail to occur, and supposing that the realisers of h are causally dependent on the realisers of m, in consequence h fails also to occur.[124]

Auch hier wird die Wahrheit des kontrafaktischen Konditionals also durch vorausgesetzte kausale Beziehungen zwischen den möglichen Basen von m und den möglichen Basen von h etabliert. In den nächsten Welten, in denen m nicht eintritt, stehen also keine Basen von m bereit, um eine Basis von h zu verursachen. Daher tritt h in diesen Welten nicht ein.

Die dargestellte Argumentation hat im Erfolgsfall unmittelbar zur Konsequenz, dass alle drei Prinzipien der Verursachung im Stufenmodell der Realität falsch sind. Denn m verursacht h, ohne p* zu verursachen. Dies widerlegt das Prinzip der Abwärtsverursachung. Zudem verursacht p p*, ohne h zu verursachen. Dies widerlegt das Prinzip der Aufwärtsverursachung. Schließlich verursacht m h, ohne dass p h verursacht. Dies widerlegt das Prinzip der basalen Verursachung. Dies legt

[123] Pernu (2016: 1043). Die Ereignis- und Satzbezeichnungen sind an die hiesige Verwendung angepasst.

[124] Pernu (2016: 1043). Die Ereignis- und Satzbezeichnungen sind an die hiesige Verwendung angepasst.

zugleich nahe, dass man im Einklang mit dieser Argumentation das Exklusionsprinzip der ontologischen Abhängigkeit akzeptieren kann. Ereignisse, die in einer ontologischen Abhängigkeitsbeziehung zueinander stehen, haben tatsächlich keine gemeinsamen Wirkungen. In diesem Sinne haben wir es hier mit einer inkompatibilistischen Erwiderung auf Exklusionsargumente zu tun: Das Exklusionsprinzip wird akzeptiert und die drei Prinzipien der Verursachung im Stufenmodell der Realität werden abgelehnt.[125]

Es ist zu beachten, dass nur die Zurückweisung des Prinzips der Abwärtsverursachung auf der Annahme einer schwach zentrierten Ähnlichkeitsrelation beruht, welche Welten, in denen m anders realisiert ist, als maximal ähnlich klassifiziert. Die Zurückweisung der anderen beiden Prinzipien beruht hingegen auf denselben Ersetzungsevaluationen der relevanten kontrafaktischen Konditionalen, die ich schon in Abschnitt 4.3.11. diskutiert habe. Ich gehe noch knapp auf die Zurückweisung des Prinzips der Abwärtsverursachung ein:

Die Begründung des Prinzips der Abwärtsverursachung in Abschnitt 4.3.11. beruhte darauf, dass aus dem gewöhnlichen kontrafaktischen Konditional ‚Wenn u nicht eingetreten wäre, dann wäre w nicht eingetreten' und dem Umstand, dass w durch eine physische Basis b necessitiert wird, folgt, dass das gewöhnliche kontrafaktische Konditional ‚Wenn u nicht eingetreten wäre, dann wäre b nicht eingetreten' wahr ist. Dies ergibt sich bereits aus der Logik kontrafaktischer Konditionale und setzt keine spezifischeren Annahmen über die Ähnlichkeitsordnung zwischen möglichen Welten voraus. Die Schlussregel, die das Prinzip der Abwärtskausalität deckt, besagt, dass aus ‚$A > B$' und ‚$\Box(B \rightarrow C)$' folgt: ‚$A > C$' und ist gültig. List und Menzies und Pernu haben hier – soweit das gewöhnliche kontrafaktische Konditional mit dem falschen Antezedens betroffen ist – auch nichts einzuwenden.[126]

Die Zurückweisung des Prinzips der Abwärtskausalität durch die Proportionalitätsforderung beruht stattdessen auf einer Betrachtung des faktischen kontrafaktischen Konditionals. Zunächst einmal ist festzustellen, dass aus dem faktischen kontrafaktischen Konditional ‚Wenn u eintreten würde, dann würde w eintreten' und der Annahme, dass w durch b necessitiert wird, nicht folgt, dass das Konditional ‚Wenn u eintreten würde, dann würde b eintreten' wahr ist. Denn diesem Schluss würde nicht die obige gültige Schlussregel zugrunde liegen, sondern die ungültige Schlussregel, der zufolge aus ‚$A > B$' und ‚$\Box(C \rightarrow B)$' folgt: ‚$A > C$'. Ein Prinzip

[125] Zugleich ist darauf hinzuweisen, dass List und Menzies (2009: 491–493, 501–502) argumentieren, dass das Exklusionsprinzip der ontologischen Abhängigkeit nicht allgemein wahr ist. Ihr Ergebnis ist jedoch damit kompatibel, dass jeder aktual auftretende Fall, in dem mentale Verursachung in Frage kommt, vom Exklusionsprinzip der ontologischen Abhängigkeit betroffen ist.

[126] Vgl. auch Pernu (2016: 1036–1037).

der Abwärtsverursachung, das auch faktische kontrafaktische Konditional berücksichtigt, kann daher nicht länger durch rein logische Überlegungen gestützt werden. Dass das Prinzip der Abwärtsverursachung nicht nur nicht durch eine logisch gültige Schlussregel gedeckt, sondern auch falsch ist, folgt dann aus der Behauptung, dass die ausschlaggebende Ähnlichkeitsrelation Realisierungsunterschiede ignorieren sollte: Denn alle maximal nahen Welten, in denen u eintritt, sind auch Welten, in denen w eintritt. Unter diesen maximal nahen Welten befinden sich jedoch auch Welten, in denen w nicht durch b sondern durch eine alternative Basis ab realisiert ist. Daher gibt es maximal nahe Welten, in denen u eintritt, b aber nicht eintritt. Also ist das Prinzip der Abwärtsverursachung falsch.

4.3.14 Kritik an der kontrafaktischen Theorie der Kausalität mit Proportionalitätsforderung

Wie überzeugend ist die im vergangenen Abschnitt rekonstruierte Argumentation für ein inkompatibilistisches Bild mentaler Verursachung?

Nach meiner Einschätzung hat die Argumentation eine Reihe von Problemen, die sie zusammengenommen eher unattraktiv erscheinen lassen. Ich führe diese Kritik im Folgenden aus:

Erstens wird die Proportionalitätsforderung hier als starke Proportionalitätsforderung in die kontrafaktische Theorie der Kausalität eingebunden: Ereignisse, die zu spezifisch oder zu unspezifisch sind, werden nicht – wie bei einer schwachen Proportionalitätsforderung – als ‘schlechtere’ Ursachen ausgezeichnet, sondern als *keine* Ursachen. Wie in Abschnitt 4.1.3. kurz angerissen, folgt hieraus, dass viele scheinbar völlig akzeptable Kausalaussagen falsch sind. Dies spricht dafür, Proportionalität nicht als notwendige Bedingung für Kausalität, sondern als Qualitätsmerkmal von Ursachen aufzufassen.

Zweitens beruft sich das Argument gegen Aufwärtsverursachung auf die fragwürdige Annahme, dass kontrafaktische Abhängigkeit notwendig für Kausalität ist. Zwar könnte man hier versuchen, die Verwendung dieser Annahme im Argument durch eine geeignete Einschränkung der Annahme (wie ‘kontrafaktische Abhängigkeit ist in geeigneten Fällen notwendig für Kausalität’) zu rechtfertigen. Es ist jedoch offen, ob man hier präziser werden kann und insbesondere die ‘geeigneten Fälle’ weiter spezifizieren kann. Wahrscheinlich kommt man um eine eingängigere Beschäftigung mit Fällen von Präemption und genuiner Überdetermination nicht herum. Die Annahme, dass kontrafaktische Abhängigkeit notwendig für Kausalität ist, spielt für die kompatibilistischen Argumente und Ergebnisse, die ich in den Abschnitten 4.3.5. bis 4.3.11 diskutiert habe, eine

weitaus weniger zentrale Rolle. Denn hier geht es in erster Linie darum, positive Aussagen über mentale Verursachung zu rechtfertigen. Hierfür kann man sich auf die Annahme, dass kontrafaktische Abhängigkeit zwischen distinkten Ereignissen hinreichend für Kausalität ist, berufen. Im Argument gegen Aufwärtsverursachung hingegen geht es darum, eine Kausalaussage auf Grundlage der Falschheit eines kontrafaktischen Konditionals zurückzuweisen. Und hierfür ist die Annahme, dass kontrafaktische Abhängigkeit notwendig für Kausalität ist, unabdingbar.

Drittens beruht das Argument gegen Aufwärtsverursachung auf einer Ersetzungsevaluation des relevanten kontrafaktischen Konditionals. Diese können jedoch – wie in Abschnitt 4.3.3. kurz angerissen – gerade in kausalen Kontexten durchaus als problematisch gesehen werden. Insbesondere stellt sich für die diskutierte Argumentation die Frage, warum nicht auch in Bezug auf das Konditional ‘Wenn m nicht eingetreten wäre, dann wäre h nicht eingetreten’ eine Ersetzungsevaluation angewendet werden sollte. Wenn in den nächsten Welten, in denen m nicht eintritt, ein sehr ähnliches Ereignis m’ eintritt, ist es durchaus denkbar, dass in diesen Welten h noch immer eintritt. Dann aber würde das Argument für höherstufige Verursachung nicht mehr funktionieren.

Viertens schließlich spielt die schwach zentrierte Ähnlichkeitsrelation, die Unterschiede mit Blick auf die Basen höherstufiger Ereignisse als irrelevante Ähnlichkeitshinsichten klassifiziert, eine tragende Rolle sowohl im Argument gegen Abwärtsverursachung als auch im Argument für höherstufige Verursachung. Auch hier kann man natürlich mit Kritik ansetzen.

Denn allgemein scheint eine schwache Zentrierungsforderung nicht wirklich gut motiviert zu sein und zu weiteren Problemen zu führen. Die grundlegende Motivation für eine schwach zentrierte Ähnlichkeitsrelation ist es, zwischen wahren und falschen faktischen kontrafaktischen Konditionalen zu unterscheiden.

Eine in Abschnitt 4.3.4. genannte Motivation, eine solche Unterscheidung zu treffen, besteht in faktischen kontrafaktischen Konditionalen, deren Antezedens irrelevant für ihr Konsequenz ist. ‘Wenn Obama 2008 zum Präsidenten der Vereinigten Staaten gewählt werden würde, dann würde die durchschnittliche Entfernung zwischen Mond und Erde 384.400 km betragen’ scheint falsch zu sein. Bei genauerem Hinsehen kann aber zumindest diese Motivation nicht überzeugen. Stattdessen können faktische kontrafaktische Konditionale sogar verwendet werden, um eben die Irrelevanz des Antezedens für das Konsequenz *auszudrücken*. ‘Wenn Obama 2008 zum Präsidenten der Vereinigten Staaten gewählt werden würde, dann würde die durchschnittliche Entfernung zwischen Mond und Erde 384.400 km betragen’ ist genauso wahr wie ‘Wenn Obama 2008 *nicht* zum Präsidenten der Vereinigten Staaten gewählt werden würde, dann würde die

durchschnittliche Entfernung zwischen Mond und Erde 384.400 km betragen'. Dass das faktische kontrafaktische Konditional ebenso wie das gewöhnliche kontrafaktische Konditional[127] mit dem negierten Antezedens wahr ist, *zeigt* gerade, dass das Antezedens irrelevant für das Konsequenz ist.

Michael McDermott nennt das folgende Beispiel eines faktischen kontrafaktischen Konditionals, das falsch zu sein scheint:

> A coin is to be tossed twice. I bet on 'Two heads', and I win. Common sense says that (1) is false. But Lewis's theory says that it is true.
>
> (1) If at least one head had come up, I would have won.[128]

Das faktische kontrafaktische Konditional 'Wenn mindestens ein Münzwurf auf Kopf gelandet wäre, dann hätte ich gewonnen' scheint falsch zu sein, obwohl tatsächlich mindestens ein Münzwurf auf Kopf gelandet ist und ich gewonnen habe. Auf den ersten Blick spricht dies für schwache Zentrierung. Denn wir haben hier ein falsches faktisches kontrafaktisches Konditional – was unter Voraussetzung starker Zentrierung ausgeschlossen ist.

Jedoch führt eine Behandlung solcher Fälle durch schwache Zentrierung zu weiteren Problemen. Um das faktische kontrafaktische Konditional 'Wenn mindestens ein Münzwurf auf Kopf gelandet wäre, dann hätte ich gewonnen' auf Grundlage einer schwach zentrierten Ähnlichkeitsrelation als falsch ausweisen zu können, muss man annehmen, dass es eine maximal ähnliche Welt gibt, in der mindestens ein Münzwurf auf Kopf gelandet ist und in der ich nicht gewinne. Plausiblerweise ist dies eine Welt, in der *genau* ein Münzwurf auf Kopf landet. Da der zweite Münzwurf in dieser Welt nicht auf Kopf landet, gewinne ich dort nicht, so dass das faktische kontrafaktische Konditional falsch ist. Es muss also die genaue Anzahl der Kopf-Münzwürfe – insbesondere der Unterschied zwischen einem und zwei Kopf-Münzwürfen – als irrelevante Ähnlichkeitshinsicht ausgezeichnet werden.

Unter Voraussetzung einer solchen Ähnlichkeitsrelation ist das faktische kontrafaktische Konditional 'Wenn ich auf zwei Kopfwürfe gewettet hätte, dann hätte ich gewonnen' jedoch *ebenfalls* falsch: Unter den maximal ähnlichen Welten finden sich schließlich noch immer auch solche Welten, in denen nur eine Münze geworfen wird. Und in diesen Welten gewinne ich nicht, obgleich ich (wie in

[127] Solche kontrafaktische Konditionale mit falschem Antezedens und wahrem Konsequenz werden gelegentlich als ‚Semi-Factuals' bezeichnet. Vgl. Walters und Williams (2013).

[128] McDermott (2007: 333).

der aktualen Welt) auf zwei Kopfwürfe gewettet habe. Das aber ist kein plausibles Ergebnis: ‘Wenn ich auf zwei Kopfwürfe gewettet hätte, dann hätte ich gewonnen’ ist intuitiv wahr – es handelt sich schließlich um ein faktisches kontrafaktisches Konditional, dessen Antezedens nicht nur wahr, sondern auch relevant für das Konsequenz ist.[129]

Man könnte diesem Einwand eventuell entweichen, indem man auf die Kontextsensitivität von kontrafaktischen Konditionalen verweist. Die unterschiedlichen kontrafaktischen Konditionale eröffnen unterschiedliche Kontexte, was sich in unterschiedlichen Ähnlichkeitsrelationen ausdrückt. Für die Bewertung von ‘Wenn ich auf zwei Kopfwürfe gewettet hätte, hätte ich gewonnen’ und ‘Wenn mindestens ein Münzwurf auf Kopf gelandet wäre, hätte ich gewonnen’ sind demnach unterschiedliche Ähnlichkeitsrelationen anzusetzen. Nur im zweiten Fall ist es angemessen, eine Ähnlichkeitsrelation anzusetzen, für die der Unterschied zwischen einem Kopf-Münzwurf und zwei Kopf-Münzwürfen irrelevant ist.

Es ist jedoch schwierig zu sehen, welche allgemeinen Überlegungen hier eine dem Kontext angemessene Ähnlichkeitsrelation festlegen könnten. Bis hierher scheint es, als würde die Ähnlichkeitsrelation im Einzelfall an den intuitiven Wahrheitswert des jeweiligen kontrafaktischen Konditionals angepasst. Dies scheint mir aber etwas zu weit zu gehen. Zwar sagt Lewis, dass “[i]t is fair to discover the appropriate standards of similarity from the counterfactuals they make true, rather than vice versa”[130]. Aber damit meint er sicher nicht, dass man die Ähnlichkeitsrelation für jedes kontrafaktische Konditional an einen unabhängig bestimmten Wahrheitswert anpassen sollte, so dass jedes kontrafaktische Konditional mit einer eigenen Ähnlichkeitsrelation im Gepäck kommt. Vielmehr sollte man die Ähnlichkeitsrelation zur möglichst allgemeinen Bewertung kontrafaktischer Konditionale aus Urteilen über eine Reihe von unterschiedlichen kontrafaktischen Konditionalen gewinnen.

Selbst wenn dieser Verlust an Allgemeinheit auf semantischer Ebene aber in Kauf genommen wird, ergeben sich aus schwacher Zentrierung weitere Probleme für die Logik kontrafaktischer Konditionale: Lee Walters argumentiert dafür, dass schwache Zentrierung und die hiermit einhergehende Ablehnung der Schlussregel der Konjunktions-Konditionalisierung, der zufolge sich aus ‘$A \wedge B$’ auf ‘$A > B$’ schließen lässt, kein trivialer, lokalisierter Eingriff in die Logik kontrafaktischer

[129] Vgl. McDermott (2007: 334).

[130] Lewis (1986b: 210–11).

Konditionale bleiben kann. Vielmehr folgt aus der Ungültigkeit dieser Schlussregel die Ungültigkeit weiterer Schlussregeln, die jedoch unabhängig plausibel sind.[131]

Diese Überlegungen legen nahe, die Proportionalitätsbedingung ohne einen Verweis auf nicht-triviale faktische kontrafaktische Konditionale zu explizieren. Und tatsächlich scheint dies auch problemlos möglich zu sein: Die Informationen, die in List und Menzies faktischem kontrafaktischen Konditional unter Voraussetzung der schwach zentrierten Ähnlichkeitsrelation stecken, können unter Voraussetzung einer stark zentrierten Ähnlichkeitsrelation durch komplexere kontrafaktische Konditionale ausgedrückt werden. Statt das faktische kontrafaktische Konditional 'Wenn m eintreten würde, dann würde h eintreten' zu betrachten und dabei davon auszugehen, dass es maximal ähnliche Welten mit anderen Basen von m (und h) gibt, kann man auch Konditionale der Art 'Wenn m eintreten würde und durch die alternative Base ap realisiert wäre, dann würde h (noch immer) eintreten' betrachten. Die Forderung, dass die Beziehung zwischen m und h realisierungsinsensitiv ist, findet ihren Ausdruck dann in der Forderung, dass solche gewöhnlichen kontrafaktischen Konditionale für jede alternative Base von m wahr sind.[132]

Das Einbinden einer schwachen Proportionalitätsforderung als Ergänzung zu den kompatibilistischen Ergebnissen der kontrafaktischen Theorie halte ich jedoch für durchaus sinnvoll. Die kontrafaktische Theorie der Kausalität ohne Proportionalitätsforderung führt zu einer bedeutenden Vervielfältigung von Kausalbeziehungen im nicht-reduktiven Physikalismus. Es ist daher wichtig, innerhalb der Vielheit der Ursachen noch ein wenig zu selektieren. Und dass mentale Ereignisse proportional zu ihren behavioralen Wirkungen sind, während ihre physischen Basen dies nicht sind, weil die Beziehung zwischen mentalen Ursachen und ihren behavioralen Wirkungen realisierungsinsensitiv ist, ist dabei ein sehr willkommenes Ergebnis. Es gesteht mentalen Ereignissen eine eingeschränkte Autonomie gegenüber ihren physischen Basen zu.

[131] Vgl. Walters (2016), vgl. auch Walters und Williams (2013).

[132] Vgl. z. B. Yablo (1992) und Harbecke (2014) für entsprechende – komplexere – Formulierungen der Proportionalitätsforderung.

4.3.15 Fazit

Ich fasse die Ergebnisse in diesem Kapitel zusammen: Die kontrafaktische Theorie der Kausalität bestimmt Kausalbeziehungen über Beziehungen kontrafaktischer Abhängigkeit. In der einfachen kontrafaktischen Theorie wird kontrafaktische Abhängigkeit zwischen gänzlich distinkten Ereignissen dabei als sowohl notwendig als auch hinreichend für Kausalität aufgefasst. Da die einfache kontrafaktische Theorie mit Problemen in Zusammenhang mit Präemption und genuiner Überdetermination konfrontiert ist, wird die Annahme, dass kontrafaktische Abhängigkeit notwendig für Kausalität ist, von vielen Autor*innen aufgegeben.

Auch von der Annahme ausgehend, dass kontrafaktische Abhängigkeit zwischen distinkten Ereignissen hinreichend für Kausalität ist, lassen sich jedoch starke Argumente für die Existenz mentaler Verursachung im nicht-reduktiven Physikalismus entwickeln. Denn die Wahrheit mentaler kontrafaktischer Konditionale im nicht-reduktiven scheint mir gut abgesichert zu sein.

Die genaue Entwicklung des Modells mentaler Verursachung, das sich aus der Anwendung der kontrafaktischen Theorie der Kausalität ergibt, ist von subtilen Fragen über die für die Auswertung der relevanten kontrafaktischen Konditionale anzusetzenden Ähnlichkeitsrelation abhängig. Akzeptiert man ein Verbot von Ersetzungsevaluation und eine stark zentrierte Ähnlichkeitsrelation, führt die kontrafaktische Theorie zu einem recht liberalen kompatibilistischen Modell mentaler Verursachung. Kausalbeziehungen vervielfältigen sich im Stufenmodell der Realität, da zumindest eingeschränkte Versionen der drei Prinzipien der Verursachung im Stufenmodell der Realität bestätigt werden.

Indem eine starke Proportionalitätsforderung in die kontrafaktische Theorie der Kausalität eingebaut wird, kann man alternativ für ein deutlich restriktiveres inkompatibilistisches Modell mentaler Verursachung argumentieren. Jedoch scheint mir diese Version der kontrafaktischen Theorie aus verschiedenen Gründen problematisch zu sein: Sie beruht auf Annahmen über die ausschlaggebende Ähnlichkeitsrelation, die letztlich nicht gänzlich überzeugen können. Zudem muss sie die Notwendigkeit kontrafaktischer Abhängigkeit für Kausalität voraussetzen und viele intuitiv einwandfreie Kausalaussagen aufgrund der starken Proportionalitätsforderung zurückweisen. Die Einbindung einer schwachen Proportionalitätsforderung in die kontrafaktische Theorie ist dennoch sinnvoll. Sie kann aufzeigen, inwiefern mentale Ereignisse als Ursachen für ihre paradigmatischen Wirkungen besser geeignet sind als physische$_{eng}$ Ereignisse. Jedoch sollte man sich für die Präzisierung einer solchen schwachen Proportionalitätsforderung

nicht auf faktische kontrafaktische Konditionale und eine schwach zentrierte Ähnlichkeitsrelation berufen, sondern auf komplexere gewöhnliche kontrafaktische Konditionale.

4.4 Die interventionistische Theorie der Kausalität und mentale Verursachung im nicht-reduktiven Physikalismus

Eine weitere Theorie der Kausalität, der in den vergangenen Jahren große Aufmerksamkeit im Zusammenhang mit dem Vorwurf des Epiphänomenalismus gegen den nicht-reduktiven Physikalismus zuteilgeworden ist, ist James Woodwards Interventionismus.

Der Interventionismus teilt mit der kontrafaktischen Theorie der Kausalität einen Fokus auf die Idee von Kausalität als Abhängigkeit: Kausalität ist demzufolge weder mit hinreichender Verursachung noch mit kausaler Produktion gleichzusetzen. Ursachen sind stattdessen Faktoren, die einen Unterschied für ihre Wirkungen machen.[133] Auch spielen – ebenso wie in der kontrafaktischen Theorie der Kausalität – kontrafaktische Konditionale eine zentrale Rolle für das interventionistische Verständnis von Kausalität. Jedoch unterscheidet sich der Interventionismus von der in Abschnitt 4.3. behandelten kontrafaktischen Theorie der Kausalität entscheidend hinsichtlich des genaueren Verständnisses kontrafaktischer Zusammenhänge. Insbesondere entfernt sich der Interventionismus von der mögliche Welten Semantik kontrafaktischer Konditionale.[134] Stattdessen werden die relevanten kontrafaktischen Konditionale unter Rückgriff auf kausale Graphen und Strukturgleichungen repräsentiert und ausgewertet.

Im Folgenden diskutiere ich die Konsequenzen des Interventionismus für mentale Verursachung im nicht-reduktiven Physikalismus. Ich gehe dabei wie folgt vor: In den Abschnitten 4.4.1. bis 4.4.3. lege ich die relevanten Grundlagen der interventionistischen Theorie der Kausalität dar. Abschnitt 4.4.4. schildert knapp die Hoffnungen einiger Autor*innen, auf der Grundlage des Interventionismus

[133] Vgl. 4.1. dieser Arbeit.

[134] Tatsächlich scheint der Interventionismus zwar kompatibel mit einer mögliche Welten Semantik zu sein, jedoch ist er nicht hierauf festgelegt. Beim gegenwärtigen Stand der Ausarbeitung der Theorie bleiben wichtige offene Fragen hinsichtlich des genauen Verständnisses der relevanten kontrafaktischen Konditionale. Siehe hierfür z. B. Hall (2007). Diese Komplikationen ignoriere ich hier, da es auch ohne eine Behandlung dieser Problematik möglich ist, die Konsequenzen des Interventionismus für den Vorwurf des Epiphänomenalismus zu diskutieren.

eine überzeugende Erwiderung auf den Vorwurf des Epiphänomenalismus zu entwickeln. Die Abschnitte 4.4.5. bis 4.4.8. dämpfen diese Hoffnung jedoch: Die Annahmen des nicht-reduktiven Physikalismus führen zu bestimmten Problemen für die interventionistische Theorie der Kausalität, die Modifikationen dieser Theorie motivieren. In den Abschnitten 4.4.9. und 4.4.10. diskutierte ich zwei Modifikationen des Interventionismus. Die erste Modifikation – der IV*-Interventionismus besteht in einer Abschwächung des Interventionismus und führt zu einer kompatibilistischen Erwiderung auf Exklusionsargumente. Die zweite Modifikation besteht in einer Verstärkung des Interventionismus durch eine starke Proportionalitätsforderung und führt zu einer inkompatibilistischen Erwiderung auf Exklusionsargumente. Ich komme auf der Grundlage dieser Diskussion zu dem Fazit (4.4.11), dass der IV*-Interventionismus in Kombination mit einer schwachen Proportionalitätsforderung zu einer zufriedenstellenden Erwiderung auf den Vorwurf des Epiphänomenalismus führt.

4.4.1 Die Grundidee des Interventionismus

Die Grundidee des Interventionismus besagt grob gesprochen, dass Verursachung konstitutiv mit bestimmten Formen der prinzipiellen Manipulierbarkeit zusammenhängt: Kausale Beziehungen sind solche Beziehungen, die im Prinzip zu Zwecken der Manipulation verwendet werden können. Wenn beispielsweise bestimmte Krankheitserreger spezifische Krankheitssymptome auslösen, so wird eine gezielte Abtötung dieser Erreger, d. h. eine Manipulation der Ursache, zum Verschwinden der Krankheitssymptome, d. h. zu einer Veränderung der Wirkung, führen. Wir können also die Wirkung manipulieren, indem wir ihre Ursache manipulieren. Laut dem Interventionismus ist es gerade diese Manipulierbarkeitsbeziehung, die die Beziehung zwischen Erreger und Symptomen zu einer kausalen Beziehung macht.[135] Woodward bringt diese Grundidee mit folgendem Slogan auf den Punkt:

> No causal difference without a difference in manipulability relations, and no difference in manipulability relations without a causal difference.[136]

[135] Vgl. Woodward (2003: 9–12).

[136] Woodward (2003: 61).

Kausalität ist dieser Grundidee zufolge also konstitutiv mit Manipulierbarkeitsbeziehungen verbunden. Was aber ist hier genau mit ‘Manipulierbarkeit’ gemeint?

Die Grundidee, Verursachung über Manipulierbarkeitsbeziehungen zu bestimmen, übernimmt der Interventionismus von früheren Manipulationstheorien der Kausalität.[137] Jedoch unterscheidet sich der Interventionismus in entscheidender Hinsicht von seinen manipulationstheoretischen Vorgängern: Er bestimmt den relevanten Sinn von Manipulation auf eine andere Weise.

Manipulation scheint alltagssprachlich zunächst an die Idee des handelnden Eingreifens gebunden zu sein. Wer manipuliert, handelt. Wenn man den Begriff der Verursachung dementsprechend durch den Begriff der Manipulation zu bestimmen versucht, bindet man den Begriff der Verursachung an den Begriff der Handlung. Einige Vertreter*innen manipulationstheoretischer Ansätze stellen diese Verbindung explizit her und verteidigen sie. So vertreten etwa Huw Price und Peter Menzies die Ansicht, dass „the ordinary notions of cause and effect have a direct and essential connection with our ability to intervene in the world *as agents*“.[138]

Manipulationstheorien, die Manipulationen unter Rückgriff auf menschliches Handeln bestimmen, sind jedoch anfällig für eine Reihe von Einwänden: Erstens laden solche Theorien dazu ein, die Frage nach der begrifflichen Bestimmung (freier) menschlicher Handlungen zu stellen. Solange dieser Begriff unbestimmt bleibt, bleibt letztlich auch die entsprechende Theorie der Kausalität unbestimmt. In anderen Worten kommt der Verdacht auf, dass hier ein unklarer Begriff durch einen wahrlich düsteren bestimmt werden soll. Zweitens wird der Begriff der Handlung häufig selbst wiederum unter Rückgriff auf kausale Begriffe bestimmt. Eine solche Bestimmung wäre jedoch zumindest dann ungeeignet, wenn der Anspruch besteht, den Kausalitätsbegriff reduktiv, d. h. *ohne* Rückgriff auf kausale Begriffe, zu bestimmen. Drittens kann gegen die begriffliche Bindung von Kausalität an menschliches Handeln eingewendet werden, dass der Kausalitätsbegriff hierdurch in dem Sinne anthropozentrisch bestimmt wird, dass er zu eng an Situationen gebunden wird, in denen Menschen handelnd eingreifen können. Kausalität ist jedoch gerade auch ein Phänomen der unbelebten Natur und findet auch dort seine Anwendung, wo handelndes Eingreifen nicht möglich ist.[139]

---

[137] Vgl. von Wright (1971) sowie Menzies und Price (1993).

[138] Vgl. Menzies und Price (1993: 187), meine Hervorhebung.

[139] Menzies und Price (1993) reagieren auf all diese Einwände. So sehen sie den Begriff der freien Handlung als hinreichend bestimmt durch direkte Bekanntschaft. Kausalität in Situationen, in denen menschliches Eingreifen nicht möglich ist, wird grob gesprochen

Woodward ist aufgrund dieser Einwände der Meinung, dass eine erfolgsversprechende Manipulationstheorie der Kausalität den zentralen Begriff der Manipulation nicht unter Bezug auf menschliches Handeln charakterisieren sollte. Stattdessen führt Woodward den technischen Begriff einer *Intervention* ein. Interventionen sind nicht dadurch ausgezeichnet, dass sie von freien Akteuren hervorgebracht werden. Vielmehr ist eine bestimmte kausale Struktur wesentlich für Interventionen. In der Bestimmung der kausalen Struktur von Interventionen orientiert sich Woodward am Aufbau randomisierter, kontrollierter Experimente:

> [I]f the reader wishes to have a concrete picture in mind of the notion of intervention that I am attempting to capture, the obvious candidate is randomized experiments.[140]

In randomisierten, medizinischen Studien zur Wirksamkeit eines Medikaments wird das Medikament nicht einfach nur an eine Reihe von Erkrankten ausgegeben. Vielmehr gibt es eine Reihe von methodologischen Standards, die eingehalten werden müssen, damit Schlüsse auf die Wirksamkeit des Medikaments zulässig sind.[141]

Ein Standard besteht darin, die erkrankten Proband*innen einer von zwei Gruppen zuzuordnen, wobei den Proband*innen in einer Gruppe das Medikament verabreicht wird, während den Proband*innen in einer Kontrollgruppe ein Placebo verabreicht wird. Die beiden Gruppen werden dann hinsichtlich des weiteren Krankheitsverlaufs verglichen. Wenn dieser Vergleich zum Ergebnis hat, dass eine Heilung in der Medikamentengruppe signifikant häufiger vorkommt als in der Kontrollgruppe, weist dies auf den ersten Blick auf eine Kausalbeziehung zwischen Medikamenteneinnahme und Heilung hin.

Jedoch müssen für einen gerechtfertigten Schluss auf eine Kausalbeziehung einige weitere, potentielle Probleme berücksichtigt werden. Zunächst einmal sollte sichergestellt werden, dass das Medikament bzw. das Placebo von den Proband*innen in der jeweiligen Gruppe auch tatsächlich eingenommen wird. Sollten die Proband*innen in der Medikamentengruppe – aus welchen Gründen auch

---

durch Ähnlichkeit zu Situationen, in denen es möglich ist, eingefangen. Inwieweit diese Erwiderungen erfolgreich sind, wird hier nicht weiter thematisiert, da lediglich Woodwards Interventionsbegriff motiviert werden soll.

[140] Woodward (2003: 95), vgl. auch Woodward (2003: 46): „Heuristically, we may think of the allowable changes in X (interventions, as we have been calling them) as processes that satisfy whatever conditions must be met in an ideal experiment designed to determine whether X causes Y.“

[141] Vgl. für die folgenden Ausführungen Woodward (2003: 94–98) und Hüttemann (2013a: 143–147).

immer – beispielsweise das Medikament nach Verabreichung regelmäßig entsorgen, ist ein Schluss auf die Wirksamkeit des Medikaments nicht mehr zulässig. Woodward verlangt daher, dass eine Intervention – in diesem Falle die Verabreichung des Medikaments – der einzige kausal relevante Faktor für die Einnahme des Medikaments ist. Das heißt, dass Faktoren, die unter normalen Umständen kausal relevant für die Einnahme des Medikaments sind (beispielsweise eine generelle Skepsis gegenüber westlicher Medizin), unter den Bedingungen des Experiments keine kausale Relevanz haben. Die Einnahme des Medikaments ist dann ganz und gar durch die Intervention bestimmt.

Aber auch, wenn sichergestellt ist, dass die Medikamentengruppe das Medikament tatsächlich einnimmt und die Placebogruppe das Placebo tatsächlich einnimmt, gibt es noch immer die Möglichkeit, dass sich der Unterschied im Krankheitsverlauf in den beiden Gruppen nicht aus der Einnahme des Medikaments ergibt, sondern auf anderen Faktoren beruht. Zum einen ist die Intervention selbst als ein solcher Faktor in Betracht zu ziehen. Sollte den verabreichten Substanzen beispielsweise anzusehen sein, ob es sich um ein Placebo oder um ein Medikament handelt, kann allein dieser Umstand zu einem Unterschied im Krankheitsverlauf führen. Woodward verlangt daher, dass die Intervention keinen Einfluss auf den Krankheitsverlauf hat, der unabhängig von der Einnahme des Medikaments ist.

Zum anderen gilt es, alle weiteren in Betracht zu ziehenden Störfaktoren auszuschließen: Die Entwicklung des Krankheitsverlaufs hat neben der Einnahme des Medikaments selbstverständlich weitere Ursachen, wie beispielsweise das Alter und die allgemeine körperliche Verfassung der Patient*innen. Um auszuschließen, dass diese anderen Faktoren den Unterschied zwischen Medikamentengruppe und Kontrollgruppe erklären, sollte die Intervention statistisch unabhängig von diesen anderen kausal relevanten Faktoren sein. Im Aufbau von Experimenten spiegelt sich dies durch die Randomisierung der Zusammensetzungen der beiden Gruppen wider. Die Probanden werden zufällig einer der beiden Gruppen zugeordnet. Bei hinreichend großen Gruppen wird so garantiert, dass sämtliche Faktoren, die sich möglicherweise unabhängig von der Medikamenteneinnahme auf den Krankheitsverlauf auswirken, in den beiden Gruppen annähernd gleich verteilt sind. Unter diesen Umständen kann der Unterschied im Krankheitsverlauf nicht auf andere Faktoren zurückgeführt werden.

Zusammenfassend verlangt Woodward also, dass eine Intervention den Faktor, auf den interveniert wird, eindeutig bestimmt, dass sie keine unabhängige Auswirkung auf die in Frage kommende Wirkung hat, und dass sie statistisch unabhängig von anderen Faktoren ist, die einen unabhängigen kausalen Einfluss auf die in Frage kommende Wirkung haben.

Woodwards Grundidee ist nun, dass der Manipulationsbegriff, der geeignet für eine Formulierung einer plausiblen Manipulationstheorie der Kausalität ist, der so bestimmte Interventionsbegriff ist. Seine interventionistische Theorie der Kausalität kann daher in einer ersten Annäherung wie folgt bestimmt werden:

> **Interventionistische Kausalität (IK):** Ein Faktor X ist genau dann eine Ursache für einen Faktor Y, wenn *eine Intervention auf X* zu einer Änderung von Y führen würde.

Statt durch menschliche Handlungen sind geeignete Manipulationen – das heißt Interventionen – also durch eine kausale Struktur gekennzeichnet, die sich am Aufbau randomisierter Experimente orientiert. Hierbei ist zu beachten, dass *jeder* Faktor als Intervention zählt, der diese kausale Struktur besitzt. Ob solche Faktoren in direktem Zusammenhang zu handelndem Eingreifen stehen oder nicht, spielt nach Woodward keine Rolle. Hierdurch geht Woodwards Interventionstheorie einigen Problemen der Theorie von Menzies und Price und anderen handlungszentrierten Manipulationstheorien aus dem Weg: Erstens befreit Woodward den Kausalitätsbegriff von der Unklarheit des Handlungsbegriffs. Zweitens ist seine Theorie nicht in dem problematischen Sinne anthropozentrisch, dass sie den Bereich kausaler Zusammenhänge auf den Bereich menschlichen Handelns einschränkt.

Dennoch ließe sich auf den ersten Blick folgender Einwand gegen Woodwards Theorie vorbringen: Woodward bestimmt den Interventionsbegriff durch eine *kausale* Struktur und verwendet diesen Begriff, um *Kausalität* begrifflich zu bestimmen. Seine Theorie ist demnach nicht reduktiv in dem Sinne, dass sie den Kausalitätsbegriff ohne Rückgriff auf kausale Begriffe bestimmt. Vielmehr muss man bereits ein Vorverständnis von Kausalität haben, um seine Definition der Kausalität zu verstehen.

Woodward reagiert auf diesen Einwand, indem er den Anspruch, eine reduktive Theorie der Kausalität zu entwickeln, ablehnt. Seine Ablehnung beruht teils auf einer generellen Skepsis gegenüber der Durchführbarkeit reduktiver Analysen philosophisch relevanter Begriffe. Zudem betont Woodward, dass eine nicht-reduktive Theorie unter bestimmten Bedingungen durchaus informativ sein und bedeutende Ergebnisse zutage fördern kann. Woodward stellt mit seiner Theorie in erster Linie Behauptungen über interessante, beidseitige, begriffliche Zusammenhänge zwischen Kausalität und Intervention auf. Diese Zusammenhänge sind auch insofern informativ, als dass sie zu Ergebnissen hinsichtlich einiger in der Philosophie der Kausalität häufig diskutierten Problemfällen führen, die von den Ergebnissen anderer Theorien abweichen und die intuitiv plausibel sind. Diese

Überlegungen führen zu der Ansicht, dass der Interventionismus zwar nicht reduktiv, jedoch auch nicht in einem problematischen Sinne zirkulär ist.

Problematisch zirkulär wäre Woodwards Theorie insbesondere dann, wenn für die Bewertung der Frage, ob eine Aussage der Art ‚Faktor X verursacht Faktor Y' wahr ist, eben jene Information über den Kausalzusammenhang *zwischen X und Y* bereits vorausgesetzt werden müsste. Es ist jedoch wichtig zu erkennen, dass Woodwards Theorie diese Struktur nicht hat. Die kausalen Informationen, die relevant für die Bewertung der Aussage ‚X verursacht Y' sind, betreffen die kausalen Beziehungen zwischen Interventionen und X sowie Informationen über andere Faktoren und Y, nicht aber den direkten Zusammenhang zwischen X und Y.[142]

Die bisherige informelle Darstellung des Interventionismus sollte die Grundidee der Theorie ausreichend spezifizieren. Sie ist jedoch in manchen Hinsichten ungenau oder vereinfachend. Das Herzstück von Woodwards Theorie besteht in einer Reihe von präziseren Definitionen unterschiedlicher kausaler Begriffe. In diesen Definitionen verwendet Woodward kausale Graphen und Strukturgleichungen als Mittel zur Repräsentation kausaler Zusammenhänge. Im folgenden Abschnitt werden zunächst einige Begriffe eingeführt, die im Zusammenhang mit kausalen Modellen stehen. Dies ermöglicht eine Erläuterung von Woodwards Definitionen im darauffolgenden Abschnitt.

4.4.2 Variablen, kausale Graphen, Strukturgleichungen und kausale Modelle

Im Rahmen des Interventionismus werden kausale Zusammenhänge als Zusammenhänge zwischen *Variablen* repräsentiert. Variablen können verschiedene *Werte* annehmen, die wiederum verschiedene Eigenschaften, Ereignisse oder Größen repräsentieren.[143] Der Wahl von Ereignissen, Eigenschaften oder Größen, die durch Werte von Variablen repräsentiert werden können, ist dabei zunächst keine Grenze gesetzt.[144] Es handelt sich bei Variablen in diesem Sinne um ein sehr flexibles Repräsentationsmittel. Für die Zwecke dieser Arbeit werde ich jedoch

[142] Vgl. Woodward (2003: 104–105).

[143] Vgl. Woodward (2003: 39).

[144] Vgl. hierzu z. B. Franklin-Hall (2016: 557): „These variables are simply anything – events, features, properties, and so on – that can vary, and are neither required to satisfy a naturalness condition, nor anything else that is 'metaphysically portentous' […]. The one restriction that variables must satisfy is that they be 'capable of taking at least two different "values"' […], meaning that they have at least two different settings or states".

in erster Linie Variablen verwenden, die das Eintreten oder Nicht-Eintreten von Ereignissen – verstanden als Eigenschaftsinstantiierungen – anzeigen.

So könnten wir zur Repräsentation der kausalen Zusammenhänge im obigen Beispiel des randomisierten Experiments etwa die in Tabelle 4.1 spezifizierten Variablen einführen.[145]

Tabelle 4.1 Variablenmenge {I, MED, H, A}

Variable	Mögliche Werte	Interpretation der Werte
I	i0, i1	i0 = Das Placebo wird verabreicht. i1 = Das Medikament wird verabreicht.
MED	med0, med1	med0 = Der Patient nimmt das Placebo ein. med1 = Der Patient nimmt das Medikament ein.
H	h0, h1	h0 = Die Krankheit des Patienten wird nicht geheilt. h1 = Die Krankheit des Patienten wird geheilt.
A	a0, a1	a0 = Der Patient ist unter 70 Jahre alt. a1 = Der Patient ist 70 Jahre alt oder älter.

Ich gehe davon aus, dass die Werte von Variablen immer in einem bestimmten Sinne *exklusiv* und *vollständig* sind. In der folgenden Passage schildert Woodward diese Forderung an Variablen:

> When considering the values of a single variable, we want those values to be logically exclusive, in the sense that variable X's taking value v excludes X's also taking value v' where v ≠ v'. We also want our variables to take a range of values corresponding to the full range of genuine or serious possibilities that can be exhibited by the system of interest.[146]

Eine Variable ist also *exklusiv*, wenn sie nicht mehrere ihrer Werte zugleich annehmen kann. Die Variable H kann beispielsweise nicht den Wert h1 und den Wert h0 zugleich annehmen: Es ist nicht möglich, dass die Krankheit sowohl geheilt als auch nicht geheilt wird. Die beiden Werte (und entsprechend die Sachverhalte, die sie repräsentieren) schließen sich gegenseitig aus. Eine Variable ist *vollständig*, wenn es nicht möglich ist, dass sie keinen ihrer Werte annimmt. Auch

[145] Ich verwende zur Veranschaulichung und zur Einführung des Vokabulars ein stark vereinfachtes und unrealistisches Beispiel. Zudem gehe ich hier und im Folgenden aus Gründen der einfacheren Darstellung davon aus, dass wir es mit deterministischen Kausalbeziehungen zu tun haben.

[146] Woodward (2016: 1064).

dieses Kriterium wird von H eindeutig erfüllt: Es ist nicht möglich, dass H weder den Wert h0 noch den Wert h1 annimmt: Die Krankheit kann nicht weder geheilt noch nicht geheilt werden.

Variable H ist offenbar exklusiv und vollständig. Das liegt einfach daran, dass ihr einer Wert die Negation ihres anderen Wertes ist. Bei der Variable MED ist die Sache hingegen weniger eindeutig. So spricht aus rein logischer Sicht natürlich nichts dagegen, dass der Patient sowohl das Placebo als auch das Medikament einnimmt und MED entsprechend den Wert med0 und den Wert med1 annimmt. Auch ist es logisch möglich, dass der Patient weder das Placebo noch das Medikament einnimmt und MED entsprechend weder den Wert med0 noch den Wert med1 annimmt. Entsprechendes gilt für Variable I.

Trotz dieser *prinzipiellen* Möglichkeiten können wir MED in bestimmten Kontexten als exklusiv und vollständig betrachten.[147] Hiermit geht jedoch eine bestimmte modale Hintergrundannahme einher: Die Möglichkeit, dass der Patient das Placebo und das Medikament bzw. weder das Placebo noch das Medikament einnimmt, kann zu Zwecken der Repräsentation der zu modellierenden kausalen Struktur ignoriert werden. Diese Möglichkeiten liegen dann in Woodwards Worten nicht im „range of genuine or serious possibilities“[148]. Mit der Wahl bestimmter Variablen geht also eine implizite Festlegung auf einen Bereich relevanter Möglichkeiten einher.[149]

Festzuhalten bleibt hier also, dass jede Variable, die ich zur Modellierung kausaler Situationen verwenden werde, als exklusiv und vollständig behandelt wird. An Stellen, an denen dies mit der impliziten Einführung eines Bereichs relevanter Möglichkeiten einhergeht, weise ich explizit darauf hin.

Die oben gewählten Variablen sind sämtlich *binär*: Sie können genau zwei Werte annehmen. Es können jedoch auch Variablen eingeführt werden, die mehr als zwei Werte haben. Veranschaulichen lässt sich dies beispielsweise an der

[147] Tatsächlich scheint mir dies auch notwendig zu sein, um größere Schwierigkeiten bei der Darstellung von kausalen Zusammenhängen zwischen Variablen zu vermeiden. Die üblichen Strukturgleichungen enthalten beispielsweise keine Informationen darüber, was mit der Wirkungsvariable passieren würde, wenn eine ihrer direkten Ursachen mehrere ihrer Werte oder keinen ihrer Werte annehmen würde.

[148] Woodward (2016: 1064), Vgl. auch Woodward (2003: 86–91) für eine Diskussion des relevanten Sinnes von „serious possibilities“.

[149] Man beachte, dass Woodward im obigen Zitat zwar die Vollständigkeit einer Variablen an einen solchen modalen Hintergrund bindet, die Exklusivität einer Variablen jedoch an Beziehungen des logischen Ausschlusses bindet. Wie das Beispiel zeigt, ist jedoch auch die Exklusivität einer Variablen an einen modalen Hintergrund gebunden.

Variable A, die das Alter des Patienten nur sehr grob anzeigt, d. h. nur repräsentiert, ob der Patient die Altersgrenze von 70 Jahren unter- oder überschreitet. Falls eine präzisere Repräsentation des Alters gewünscht ist, könnte man alternativ zum Beispiel eine Variable H_{mult} einführen, die den Wert $h_{mult}1$ annimmt, wenn der Patient 1 Jahr alt ist, den Wert $h_{mult}2$ annimmt, wenn der Patient 2 Jahre alt ist, usw. Eine solche Variable hätte nicht nur zwei, sondern zahlreiche mögliche Werte. Weiter unten werde ich u. a. auch von solchen *mehrwertigen* Variablen Gebrauch machen.

Bleiben wir zur Illustration jedoch bei der oben eingeführten Variablenwahl und der Variablenmenge {I, MED, H, A}. Ein zentrales Mittel zur Repräsentation kausaler Beziehungen im Rahmen des Interventionismus besteht in *kausalen Graphen.*[150] In kausalen Graphen werden kausale Beziehungen zwischen Variablen repräsentiert. Um die kausalen Beziehungen zwischen I, MED, H und A zu repräsentieren, könnten wir beispielsweise den kausalen Graphen in Abbildung 4.3 verwenden:

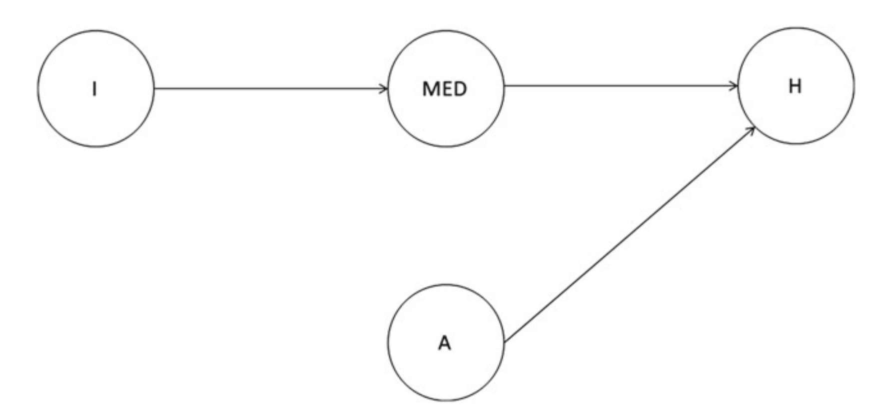


Abbildung 4.3 Ein einfacher kausaler Graph für die Variablenmenge {I, MED, H, A}

Ein Pfeil von einer Variable X zu einer anderen Variable Y in einem solchen kausalen Graphen bedeutet, dass X eine *direkte Ursache* für Y ist. Die Variable H, die die Heilung eines Patienten repräsentiert, hat demnach zwei direkte Ursachen: Die Variable MED, die die Einnahme des Medikaments repräsentiert, und die Variable A, die das Alter des Patienten repräsentiert. Variable MED wiederum

[150] Vgl. für das folgende Woodward (2003: 42), Hüttemann (2013: 147–154) oder Hitchcock (2018).

hat eine direkte Ursache – Variable I, die die Verabreichung des Medikaments repräsentiert.

I ist keine *direkte* Ursache von H, da es keinen Pfeil von I zu H gibt. Jedoch gibt es einen *kausalen Pfad* von I zu H, weil I eine direkte Ursache von MED und MED eine direkte Ursache von H ist. Im Allgemeinen besteht (informell) genau dann ein kausaler Pfad von einer Variable X zu einer Variablen Y, wenn es eine Kette aus direkten Ursachen von X zu Y gibt, wobei die Kette auch aus nur den zwei Gliedern X und Y bestehen kann.

Die Verhältnisse zwischen den Variablen in einem kausalen Graphen werden oft durch genealogisches Vokabular erläutert: Wenn es einen kausalen Pfad von einer Variable X zu einer Variablen Y gibt, heißt X ein *Vorfahre* von Y und Y ein *Nachfahre* von X. I ist also Vorfahre von MED und von H; und H ist Nachfahre von MED, von I und von A. Die direkten Ursachen einer Variable X werden auch *Eltern* von X genannt. H hat im obigen Graphen also die Eltern MED und A. *Endogene Variablen* in einem Graphen werden Variablen genannt, die relativ zu dem Graphen Eltern haben. *Exogene Variablen* in einem Graphen sind dagegen solche, die relativ zu dem Graphen keine Eltern haben. I und A sind im gegebenen Graphen also exogene Variablen, während MED und H endogene Variablen sind.

Kausale Graphen sind gut dazu geeignet, zu repräsentieren, zwischen welchen Variablen in einer gegebenen Variablenmenge überhaupt kausale Beziehungen bestehen. Jedoch enthalten sie noch keine Informationen darüber, auf welche Weise sich eine Wirkungsvariable nach Veränderungen ihrer Eltern verändert. So ist der kausale Graph in Abbildung 4.3 beispielsweise zunächst sowohl damit kompatibel, dass die Heilung der Krankheit (H = h1) nur dann eintritt, wenn der Patient über 70 Jahre alt (A = a1) ist und das Placebo einnimmt (MED = med0), als auch damit, dass die Heilung nur dann eintritt (H = h1), wenn der Patient unter 70 Jahre alt ist (A = a0) und das Medikament einnimmt (MED = m1). Der kausale Graph legt bloß fest, dass die Heilung von den Faktoren ‚Alter' und ‚Medikamenteneinnahme' abhängt, aber nicht, wie diese Abhängigkeit genauer beschaffen ist.

Um die genaueren funktionalen Zusammenhänge zwischen unterschiedlichen Variablen zu spezifizieren, werden *Strukturgleichungen* verwendet. Für jede endogene Variable gibt es eine *Strukturgleichung*, die spezifische Informationen darüber enthält, wie sich der Wert der Variable aus den Werten ihrer Eltern ergibt. Der Wert von H im obigen Modell ist demnach eine Funktion aus den Werten von MED und A – also der Eltern von H:

(SG1) $H := F_H(MED,A)$

(SG1) weist jeder Kombination der Werte von MED und A eindeutig einen Wert von H zu. Gelegentlich wird es sich als sinnvoll erweisen, die strukturellen Zusammenhänge zwischen Variablen möglichst exakt darzustellen. Eine einfache Darstellungsmöglichkeit besteht in Funktionstabellen, die jeder Kombination der Werte der Eltern einer endogenen Variablen eindeutig einen Wert der jeweiligen Funktion und somit einen Wert der in Frage stehenden Variable zuordnen. Ein Beispiel findet sich in Tabelle 4.2.

Tabelle 4.2 Funktionstabelle für Gleichung H := F_H(MED,A)

MED	A	F_H(MED,A)
med0	a0	h0
med0	a1	h0
med1	a0	h1
med1	a1	h0

Die hiermit ausgedrückte Strukturgleichung ‚H: = F_H(MED,A)‘ würde entsprechend bedeuten, dass sich eine Heilung nach Einnahme des Medikaments nur bei unter 70-Jährigen einstellt.

Ein *kausales Modell* schließlich besteht aus einer Variablenmenge und einer Menge von Strukturgleichungen, die die Zusammenhänge zwischen den Variablen in dieser Variablenmenge spezifizieren. So kann ein kausales Modell, das die Abhängigkeiten zwischen MED, A und H, spezifiziert, beispielsweise als ein Paar aus der Variablenmenge {MED, A, H} und der Menge der Strukturgleichungen {F_H(MED,A)} verstanden werden – wobei die Menge der Strukturgleichungen in diesem einfachen Fall nur ein Element hat.

4.4.3 Intervention und Verursachung – Definition zentraler Begriffe

Woodward unterscheidet eine Reihe von unterschiedlichen kausalen Begriffen wie ‚totale Ursache‘, ‚direkte Ursache‘, ‚beitragende Ursache‘ und ‚aktuale Ursache‘. All diese Begriffe werden durch ihre Beziehung zu möglichen Interventionen bestimmt – allerdings auf unterschiedliche Weise. Alle Begriffe werden zudem unter Verwendung des im vorigen Abschnitt eingeführten Vokabulars ausgedrückt. Auch der Begriff einer Intervention selbst wird noch einmal unter Verwendung dieses Vokabular präzisiert. Im Folgenden werde ich zunächst Woodwards Definition einer Intervention erläutern, um daraufhin darauf einzugehen,

wie Woodward die angesprochenen kausalen Begriffe in Bezug auf Interventionen definiert.

Wie oben gesehen verlangt Woodward von einer Intervention, dass sie den Faktor, auf den interveniert wird, eindeutig bestimmt, dass sie keine unabhängigen Auswirkungen auf den Faktor hat, der als Ursache in Frage kommt, und dass sie statistisch unabhängig von anderen Faktoren ist, die Ursachen sind für den Faktor, der als Ursache in Frage kommt. Unter Verwendung des Vokabulars kausaler Graphen fängt Woodward diese Bedingungen mit folgenden Definitionen ein.

Zunächst definiert er den Begriff eine *Interventionsvariablen*:

> **(IV)** Eine Variable I ist eine Interventionsvariable auf eine Variable X hinsichtlich einer Variable Y genau dann, wenn gilt:
>
> (I1) I verursacht X.
>
> (I2) I ist ein Umschalter für X. Das heißt, bestimmte Werte von I sind so beschaffen, dass gilt: Wenn I diese Werte annimmt, hängt X nicht von den Werten anderer Ursachen von X ab, sondern hängt stattdessen nur vom Wert von I ab.
>
> (I3) Jeder kausale Pfad von I zu Y geht durch X. Das heißt, I ist keine direkte Ursache von Y und ist keine Ursache von Ursachen von Y, die distinkt von X sind, außer jenen Ursachen von Y, die in den gerichteten Pfad von I über X zu Y eingebaut sind; also (a) jenen Ursachen von Y, die Wirkungen von X sind und (b) jenen Ursachen von Y, die auf einem kausalen Pfad zwischen I und X liegen und die keine von X unabhängigen Auswirkungen auf Y haben.
>
> (I4) I ist statistisch unabhängig von jeder Variablen Z, die Y verursacht und die auf einem kausalen Pfad liegt, der nicht durch X geht.[151]

(IV) definiert den Begriff einer Interventionsvariablen auf X *hinsichtlich Y*. Der Begriff einer Interventionsvariablen auf X ist mithin relativiert auf eine weitere Variable. Bei dieser Variable Y handelt es sich um die Variable, die als Wirkung von X in Frage kommt. Im obigen Beispielgraphen wäre dies also Variable H, die die Heilung der Patienten repräsentiert.

Bedingungen (I1) und (I2) sollen dafür sorgen, dass Interventionen den Wert der Variable X, auf die interveniert wird, vollständig bestimmen. Sie stellen daher Forderungen an die kausalen Beziehungen, in denen I und X stehen. Variable Y

[151] Vgl. Woodward (2003: 98).

spielt hier jedoch noch keine Rolle. (I1) verlangt einfach, dass eine Interventionsvariable I auf X eine Ursache für X ist. (I2) verstärkt diese Forderung: I soll nicht nur irgendeine Ursache für X sein, sondern ein ‚Umschalter' für X.[152]

Der in (I2) vorkommende Begriff eines Umschalters lässt sich wie folgt veranschaulichen:[153] In einem Audio-System sind verschiedene Faktoren relevant für den Klang, der aus den Lautsprechern kommt. So kann man den Klang etwa durch eine Einstellung des Lautstärkereglers, durch eine Einstellung des Bassreglers und durch eine Einstellung des Höhenreglers am Verstärker verändern. All diese Faktoren sind Ursachen für die Klangausgabe. All diese Faktoren machen jedoch keinen Unterschied mehr, sobald der Verstärker ausgeschaltet ist. Wenn wir also die Position des Einschaltknopfs des Verstärkers durch eine Variable E repräsentieren, die den Wert 0 annimmt, wenn der Verstärker ausgeschaltet ist und den Wert 1, wenn er eingeschaltet ist, dann ist E ein Umschalter für die Klangausgabe im relevanten Sinn: Der Wert 0 von E sorgt dafür, dass die Abhängigkeitsbeziehungen zwischen der Klangausgabe und der Einstellung der anderen Regler gebrochen werden. Wenn E den Wert 0 hat, legt dies alleine bereits vollkommen fest, dass kein Klang aus den Lautsprechern kommt. Die Einstellung der anderen Regler ist dann irrelevant. In diesem Sinne ist auch die Verabreichung des Medikaments im obigen Beispiel (I) im Idealfall ein Umschalter für die Medikamenteneinnahme (MED). Die Faktoren, die unter gewöhnlichen Umständen eine Rolle dafür spielen würden, ob die Patient*innen das Medikament einnehmen, sind – gegeben, dass das Medikament unter den kontrollierten Bedingungen des Experiments verabreicht wird – irrelevant. Wenn I also den Wert 1 annimmt, legt dies alleine bereits fest, dass auch MED den Wert 1 annimmt.

Bedingungen (I3) und (I4) sollen dafür sorgen, dass eine etwaige Änderung von Y nach Intervention auf X nicht auf Faktoren zurückzuführen ist, die von X unabhängig sind. Dabei kämen zum einen ‚Nebenwirkungen' der Intervention selbst in Frage. Wenn die Intervention I die mögliche Wirkung Y direkt beeinflusst oder auf einer Art und Weise indirekt beeinflusst, die X nicht mit einbezieht, könnte eine etwaige Änderung von Y nach Intervention auf X auch durch diese Wirkung von I auf Y erklärt werden. Y könnte sich dann durch die ‚Intervention' ändern, ohne dass X dafür eine Rolle spielt. Dies wird durch (I3) ausgeschlossen.

(I4) soll sämtliche Störfaktoren ausschließen, die nicht von I verursacht werden. Als Beispiel kommt mit Blick auf die Medikamentenstudie das Alter der Patient*innen in Betracht. Die Intervention sollte statistisch unabhängig vom

[152] Entsprechend ist (I1) eigentlich redundant. Ich bleibe dennoch bei der vorliegenden Definition, weil sie in der zu behandelnden Literatur in aller Regel vorausgesetzt wird.
[153] Vgl. Woodward (2003: 96–97).

Alter der Patient*innen (A) sein, damit etwaige Unterschiede im Heilungsverlauf nicht auf Unterschiede im Alter zurückgeführt werden können. Im Allgemeinen sollte I statistisch unabhängig von allen Ursachen der als Wirkung in Frage kommenden Variable Y sein.[154]

Gegeben diese Definition einer Interventionsvariablen auf X hinsichtlich Y definiert Woodward den Begriff einer (tatsächlichen) *Intervention* nun wie folgt:

> **(IN)** Dass I einen Wert $I=z_i$ annimmt, ist eine Intervention auf X hinsichtlich Y, genau dann wenn gilt: $I=z_i$ ist eine aktuale Ursache für den Wert, den X annimmt.[155]

Eine tatsächliche Intervention ist also ein Wert einer Interventionsvariablen I, der den Wert von X aktual verursacht.

Sowohl in (IV) als auch in (IN) beruft sich Woodward auf kausale Begriffe wie ‚direkte Ursache' und ‚aktuale Ursache', die bis hierher undefiniert sind. Woodward definiert diese Begriffe nun unter Rückgriff auf Interventionen. Sein System von Definitionen bildet also ein Paket, in welchem die verschiedenen Definitionen wechselseitig aufeinander bezogen sind. Zur weiteren Erläuterung von (IN) und (IV) ist es dementsprechend nötig, die verwendeten kausalen Begriffe einzuführen.

Als Ausgangspunkt von Woodwards Definitionen kausaler Begriffe kann die unter 4.4.1. angegebene erste Annäherung dienen:

> **Interventionistische Kausalität (IK):** Ein Faktor X ist genau dann eine Ursache für einen Faktor Y, wenn eine Intervention auf X hinsichtlich Y zu einer Änderung von Y führen würde.

Wenn diese Ausgangsidee relativ direkt auf das eingeführte Vokabular übertragen wird, ergibt sich Woodwards Definition einer totalen Ursache:

---

[154] Auch der Fall, in dem eine Ursache von I einen unabhängigen Einfluss auf Y hat, wird durch (I4) ausgeschlossen, da I von ihren Ursachen nicht statistisch unabhängig ist. Vgl. Kaiserman (2018).

[155] Vgl. Woodward (2003: 98). Hier ist, soweit ich sehe, eine kleine Ungenauigkeit in Woodwards Definitionen versteckt: Denn (IV) lässt zu, dass I auch Werte hat, die den Einfluss anderer Faktoren auf X *nicht* abschneiden. (IN) lässt zu, dass diese Werte den Wert von X aktual verursachen und dennoch als Interventionen zählen. Dann aber kann der Wert von X unter einer Intervention von anderen Ursachen mitbestimmt werden. Es müsste eigentlich in (IN) verlangt werden, dass es sich bei Interventionen um Umschalter-Werte für I handelt. Alternativ könnte in (IV) verlangt werden, dass *jeder* Wert von I ein Umschalter-Wert ist.

Totale Ursache (TU): Eine Variable X ist eine totale Ursache für eine Variable Y, genau dann, wenn gilt: Es gibt eine mögliche Intervention auf X hinsichtlich Y, die Y (oder die Wahrscheinlichkeitsverteilung für Y) verändern würde.[156]

Diese Definition (wie auch Woodwards weitere Definitionen) verlangt nicht, dass *jede* Intervention auf X hinsichtlich Y zu einer Änderung von Y führt. Stattdessen muss es nur *eine* Intervention auf X hinsichtlich Y geben, die Y verändern würde. Hieraus folgt, dass mehrere Werte von X demselben Wert von Y zugeordnet sein können. Beispielsweise kann X auch dann eine (totale) Ursache von Y sein, wenn eine Intervention, die X auf den Wert × 0 setzt, zu keiner Änderung von Y führt, solange eine andere Intervention, die X etwa auf den Wert × 2 setzt, zu einer Änderung von Y führt. Entsprechende Zusammenhänge finden sich beispielsweise, wenn eine Variable X einen Schwellenwerteffekt auf Y ausübt, so dass X einen bestimmten Wert überschreiten muss, damit sich Y ändert.

Laut Woodward ist totale Verursachung hinreichend[157], aber nicht notwendig für Verursachung. Das heißt: Jede totale Ursache ist eine Ursache. Jedoch gibt es Ursachen, die keine totalen Ursachen sind.

Der Begriff einer totalen Ursache ist daher alleine nicht dafür geeignet, die kausalen Strukturen, die in kausalen Graphen ausgedrückt werden, einzufangen. Denn es ist nicht wahr, dass ein Pfeil in einem kausalen Graphen von X zu Y genau dann führt, wenn X eine totale Ursache für Y ist. Vielmehr kann X auch dann eine direkte Ursache für Y sein, wenn *keine* mögliche Intervention auf X Y verändern würde. Dies ist unter anderem dann der Fall, wenn zwei gerichtete Pfade von X zu Y führen, die sich exakt ausgleichen.

Christopher Hitchcock gibt hierfür folgendes Beispiel:[158] Man betrachte die historische Hypothese, dass der Bau von Eisenbahnen in den Vereinigten Staaten des 19. Jahrhunderts ein ausschlaggebender Faktor für gesunkene Transportkosten war. Gegen diese Hypothese wurde eingewendet, dass ohne den Bau von Eisenbahnen mehr Kapital in die Entwicklung und den Ausbau alternativer Transportmittel gesteckt worden wäre, was wiederum die Transportkosten reduziert hätte. Die Idee ist demnach, dass der Bau der Eisenbahnen zwar durchaus die Transportkosten reduziert hat, dass sie aber zugleich die Entwicklung alternativer Transportmittel wie Straßen und Lastwagen verhindert oder verlangsamt hat, die

[156] Vgl. Woodward (2003: 51). Da ich die Diskussion im Folgenden auf Fälle deterministischer Kausalität beschränke, kann die Klausel ‘oder die Wahrscheinlichkeitsverteilung für Y’ ignoriert werden.

[157] Vgl. Woodward (2003: 49).

[158] Vgl. Hitchcock (2001: 364). Vgl. auch Woodward (2003: 50–53), der ein strukturell gleiches Beispiel verwendet.

in Abwesenheit des Eisenbahnbaus ebenso zu einer Senkung der Transportkosten geführt hätte. Eine Intervention ausschließlich auf den Bau der Eisenbahnen hätte daher (so nehmen wir idealisierend an) nicht zu höheren Transportkosten geführt, sondern dazu, dass die Transportkosten auf anderem Wege in gleichem Maße reduziert worden wären. Daher kann der Bau der Eisenbahnen nicht als totale Ursache für die gesunkenen Transportkosten im Sinne von (TU) gelten: Eine isolierte Intervention auf diesen Faktor hätte die Transportkosten nicht verändert. Dennoch scheint es sich in einem klaren Sinn um eine Ursache zu handeln. Die kausale Struktur kann durch den Graphen in Abbildung 4.4 dargestellt werden:

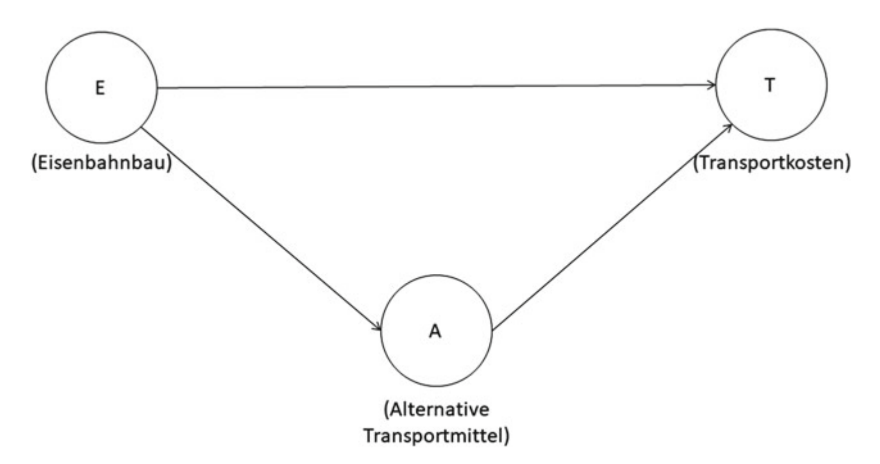


Abbildung 4.4 kausaler Graph für die Variablenmenge {E, A, T}

Der Pfeil von der Eisenbahnbau-Variablen E zur Transportkosten-Variable T kann hier nicht bedeuten, dass eine isolierte Intervention auf den Eisenbahnbau zu einer Änderung der Transportkosten-Variable führen würde. Denn eine solche Intervention auf E würde über den gerichteten Pfad von E über A zu T dazu führen, dass die Transportkosten in demselben Maße gesenkt werden wie ohne die Intervention. Die Definition einer Intervention verlangt nicht, dass eine Intervention auf E hinsichtlich T statistisch oder kausal unabhängig von A ist, weil A T nicht unabhängig von E beeinflusst, sondern auf einem kausalen Pfad liegt, der E einschließt (siehe (I4)). Daher würde jede Intervention auf E auch A auf eine Weise beeinflussen, die den Effekt von E auf T ausgleicht, so dass sich T bei Interventionen auf E nicht verändert.

Zugleich ist jedoch zu beachten, dass eine Intervention auf E zu einer Änderung von T führen würde, *wenn A auf einem bestimmten Wert festgehalten würde*: Wenn eine zusätzliche Intervention auf A dafür sorgt, dass keine alternativen Transportmittel entwickelt werden, dann senkt eine Intervention auf E, die dafür sorgt, dass Eisenbahnen gebaut werden, die Transportkosten (T). Es scheint dieser Zusammenhang zwischen E, A und T zu sein, der begründet, dass ein Pfeil von E zu T gezogen werden sollte, obgleich E keine totale Ursache für T ist.

Woodward fängt diese Idee mit seiner Definition der direkten Verursachung ein:

> **Direkte Ursache (DU):** Eine Variable *X* ist eine direkte Ursache einer Variablen *Y* relativ zu einer Variablenmenge V genau dann, wenn gilt: Es gibt eine mögliche Intervention auf *X* hinsichtlich *Y*, die *Y* (oder die Wahrscheinlichkeitsverteilung für Y) verändert, *wenn alle anderen Variablen in V außer X und Y durch Interventionen auf einem Wert festgesetzt werden.*[159]

Gegeben diese Definition ist E aufgrund der obigen Überlegung also eine direkte Ursache für T: Wenn A auf einem Wert festgehalten wird, ändert eine Intervention auf E auch T. Drei Kommentare zu dieser Definition:

Erstens: Pfeile in kausalen Graphen werden an diesen Begriff der direkten Verursachung gebunden: Ein Pfeil von einer Variable X zu einer Variable Y darf in einem kausalen Graph genau dann gezogen werden, wenn X eine direkte Ursache für Y ist. Hierdurch nimmt der Begriff der direkten Verursachung eine zentrale Rolle innerhalb von Woodwards System kausaler Begriffe ein: Die weiteren Begriffe verweisen alle auf die eine oder andere Weise auf direkte Verursachung.

Zweitens: Wie auch in (TU) wird in (DU) nicht verlangt, dass *jede* Intervention auf X zu einer Änderung von Y führt, wenn die anderen Variablen auf einem Wert festgehalten werden, sondern lediglich, dass es eine solche Intervention gibt. Zudem wird in (DU) in ähnlicher Weise nicht verlangt, dass eine Intervention auf X Y unter *allen möglichen* Kombinationen von Interventionen auf die anderen Variablen verändert, sondern lediglich, dass es *eine* solche Kombination von Interventionen auf die anderen Variablen gibt, unter der eine Intervention auf X Y verändert. Bezogen auf das obige Beispiel heißt das, dass der Eisenbahnbau auch dann als direkte Ursache für die Transportkosten gelten kann, wenn eine Intervention, die dafür sorgt, dass *in jedem Fall* alternative Transportmittel entwickelt werden, zur Konsequenz hat, dass eine Intervention auf den Eisenbahnbau keinen zusätzlichen Unterschied für die Transportkosten mehr macht. Solange der Eisenbahnbau einen Unterschied macht, wenn eine Intervention dafür sorgt, dass die

[159] Vgl. Woodward (2003: 55).

alternativen Transportmittel *nicht* entwickelt werden, kann der Eisenbahnbau als direkte Ursache gelten. Denn dann gibt es zumindest *eine* Intervention auf die alternativen Transportmittel, unter der eine Intervention auf den Eisenbahnbau eine Änderung der Transportkosten zur Folge hat.

Drittens: (DU) bestimmt den Begriff der direkten Ursache *relativ zu einer Variablenmenge V*. Es stellt sich die Frage, welche Konsequenzen sich aus einer solchen Relativierung des Begriffs der direkten Verursachung auf eine Variablenmenge ergeben. Insbesondere stellt sich die Frage, ob eine Variable X direkte Ursache für Y relativ zu einer Variablenmenge V sein kann, ohne direkte Ursache für Y relativ zu einer anderen Variablenmenge V‘ zu sein, die X und Y ebenfalls enthält. Ich gehe auf diese Frage weiter unten noch einmal ausführlicher ein. Bereits jetzt sind jedoch zwei Anmerkungen hierzu angebracht:

Erstens kann es vorkommen, dass X keine direkte Ursache für Y relativ zu einer Variablenmenge V ist, aber X eine direkte Ursache für Y relativ zu V + ist, wobei V + eine Erweiterung von V darstellt.[160] Das heißt: Pfeile zwischen Variablen in einem kausalen Graphen können auftauchen, wenn weitere Variablen hinzugefügt werden. Hitchcocks Beispiel des Eisenbahnbaus vermag dies zu veranschaulichen: Betrachten wir hier zunächst die Variablenmenge {E, T}, die also nur die Variablen für den Eisenbahnbau und die Transportkosten enthält. E wird sich nicht als direkte Ursache für T relativ zu {E, T} qualifizieren. Denn die Definition der direkten Verursachung verlangt nicht, dass wir A (Entwicklung alternativer Transportmittel) festhalten, wenn es um die Frage geht, ob E eine direkte Ursache für T *relativ zu {E, T}* ist. Der Grund hierfür ist einfach, dass A nicht in der Variablenmenge {E, T} enthalten ist. Die Definition der direkten Verursachung verlangt jedoch nur, dass alle Variablen in der jeweiligen Variablenmenge festgehalten werden. Sie verlangt entsprechend nicht, dass A festgehalten wird. Tatsächlich verlangt die Definition der direkten Verursachung mit Blick auf die Variablenmenge {E, T} nicht mehr als die Definition der totalen Verursachung, die Woodward ja aufgrund dieses Beispiels modifiziert: Da eine Intervention auf E zu einer Änderung von A führt, die wiederum den Effekt von E auf T ausgleicht, wird sich T wie oben geschildert bei einer isolierten Intervention auf E nicht verändern. E ist also keine direkte Ursache für T relativ zu {E, T}.

Betrachten wir nun eine Erweiterung von {E, T}, nämlich {E, T, A}. Relativ zu {E, T, A} wird sich E nun als direkte Ursache für T herausstellen. Denn es

[160] Eine Erweiterung V + einer Variablenmenge V ist V + genau dann, wenn V eine echte Teilmenge von V + ist.

gibt eine Intervention auf E, die T verändert, wenn A auf einem Wert festgehalten wird. Und da A nun Teil der relevanten Variablenmenge ist, verlangt die Definition der direkten Ursache, dass A auf einem Wert festgehalten wird. Die Definition der direkten Verursachung hat also nur das gewünschte Ergebnis, dass E eine direkte Ursache für T ist, wenn eine hinreichend reichhaltige Variablenmenge betrachtet wird: Der ausgleichende Pfad von E zu T über A muss in der relevanten Variablenmenge repräsentiert sein.[161]

Zweitens kann es aus einem recht einfachen Grund vorkommen, dass eine Variable X eine direkte Ursache für eine Variable Y relativ zu einer Variablenmenge V ist, ohne dass X eine direkte Ursache für Y relativ zu einer Erweiterung V + von V ist. Dies wird immer dann der Fall sein, wenn V + eine zusätzliche Variable Z enthält, die auf einem kausalen Pfad zwischen X und Y liegt.[162] Zur Veranschaulichung greife ich wieder auf das Beispiel der medizinischen Wirksamkeitsstudie zurück: Relativ zur Variablenmenge {MED, H}, die nur Variablen für die Medikamenteneinnahme und die Heilung der Krankheit enthält, ist MED eine direkte Ursache für H: Eine Intervention auf MED würde H verändern, während alle anderen Variablen in der Variablenmenge (in diesem Fall keine) festgehalten werden. Relativ zu einer Variablenmenge, die genauer auf den Mechanismus eingeht, der von MED zu H führt, und zum Beispiel eine zusätzliche Zwischenvariable enthält, die die Konzentration bestimmter Viren im Blut (VK) repräsentiert, zählt MED jedoch nicht als direkte Ursache für H (siehe Abbildungen 4.5 und 4.6).



Abbildung 4.5 kausaler Graph für die Variablenmenge {MED, H}

Wenn alle anderen Variablen außer MED und H in der Variablenmenge {MED, VK, H} auf einem Wert festgehalten werden, ändert eine Intervention auf MED

[161] Vgl. Woodward (2003: 56). In diesem Beispiel haben wir es mit nur einem ausgleichenden Pfad von E zu T zu tun. Es sind jedoch auch Beispiele denkbar, in denen zahlreiche ausgleichende Pfade im Hintergrund stehen (vgl. hierzu auch Woodward (2003: 54)). In solchen Fällen muss die betrachtete Variablenmenge noch reichhaltiger sein: Sie muss in jedem Falle alle ausgleichenden Pfade repräsentieren.

[162] Vgl. Woodward (2003: 55–56).

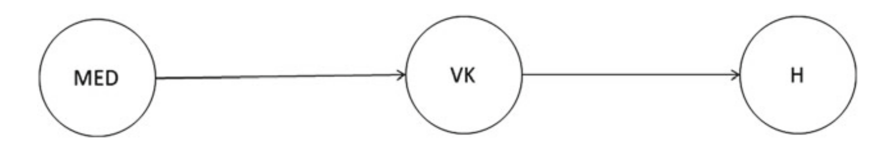


Abbildung 4.6 kausaler Graph für die Variablenmenge {MED, VK, H}

nicht H. Denn wenn VK auf einem Wert festgehalten wird – d. h. die Konzentration der Viren im Blut aufgrund einer weiteren Intervention auch bei Einnahme des Medikaments unverändert bleibt – führt eine Einnahme des Medikaments nicht zur Heilung der Krankheit. MED ist also keine direkte Ursache für H relativ zu {MED, VK, H}.

Offenbar können wir MED dennoch in einem gewissen Sinne als Ursache von H relativ zu {MED, VK, H} ansprechen. Schließlich ist MED eine direkte Ursache für VK und VK eine direkte Ursache für H. Um diesen Sinn von Verursachung einzufangen, führt Woodward den Begriff der beitragenden Verursachung ein:[163]

> **Beitragende Ursache (BU):** Eine Variable X ist eine beitragende Ursache einer Variablen Y relativ zu einer Variablenmenge V genau dann, wenn gilt: (i) Es gibt einen kausalen Pfad P, d.h. eine Kette aus direkten Ursache, der von X zu Y führt und (ii) es gibt eine Intervention auf X, die Y (oder die Wahrscheinlichkeitsverteilung für Y) verändert, wenn alle Variablen in V, die nicht auf dem kausalen Pfad P von X zu Y liegen, auf einem Wert festgehalten werden.[164]

Der Unterschied zwischen beitragenden und direkten Ursachen besteht im Wesentlichen darin, dass beitragende Ursachen Zwischenschritte erlauben. Der Pfad von X zu Y kann also weitere Variablen enthalten. Natürlich sollten solche Variablen, die auf einem kausalen Pfad zwischen X und Y liegen, nicht durch Interventionen auf einem Wert festgehalten werden. Denn solche Interventionen würden dafür sorgen, dass die Zwischenvariablen unabhängig von X werden und

---

[163] Woodward wählt hier nicht einfach die transitive Hülle der Relation der direkten Verursachung, weil er aufgrund von Gegenbeispielen gegen die Transitivität von Verursachung meint, dass beitragende Verursachung nicht als transitive Relation verstanden werden sollte (vgl. Woodward (2003: 57–59)). Da die Frage nach der Transitivität der Kausalrelation keine direkten Konsequenzen für das hier diskutierte Problem hat, übergehe ich Woodwards genauere Begründung an dieser Stelle.

[164] Vgl. Woodward (2003: 59).

der Einfluss von X auf Y daher nicht mehr über die Zwischenvariablen übertragen wird. Es wird daher in (BU) leidglich verlangt, dass die anderen Variablen in der Variablenmenge, die *nicht* auf dem kausalen Pfad von X und Y liegen, auf einem Wert festgehalten werden. Entsprechend ist MED in Abbildung 4.6 eine beitragende Ursache für H relativ zu {MED, VK, H}. Denn es wird hier nicht verlangt, dass VK auf einem Wert festgehalten wird.

Auch (BU) bestimmt einen auf Variablenmengen relativierten Begriff beitragender Verursachung. Und auch mit Blick auf beitragende Verursachung kann es vorkommen, dass eine Variable X keine beitragende Ursache für eine Variable Y relativ zu einer Variablenmenge V ist, aber X eine beitragende Ursache für Y relativ zu einer Erweiterung V + von V ist. Der Grund dafür ist derselbe wie bei direkter Verursachung: Wenn ausgleichende Pfade in V nicht repräsentiert werden, verlangt (BU) nicht, dass die Variablen auf den ausgleichenden Pfaden festgehalten werden. Fügt man Variablen auf den ausgleichenden Pfaden aber hinzu, müssen diese festgehalten werden. E ist im Eisenbahnbeispiel entsprechend keine beitragende Ursache relativ zu {E, T}, aber wird zu einer beitragenden Ursache relativ zu {E, A, T}.

Anders als im Fall der direkten Verursachung können Beziehungen der beitragenden Verursachung laut Woodward jedoch nicht verschwinden, wenn weitere Variablen zu einer Variablenmenge hinzugefügt werden: Wenn X eine beitragende Ursache für Y relativ zu einer Variablenmenge V ist, dann ist X auch eine beitragende Ursache für Y relativ zu jeder Erweiterung V + von V. In diesem Sinne ist die Relation der beitragenden Verursachung *monoton.*[165] Zudem ist jede direkte Ursache relativ zu einer Variablenmenge V auch eine beitragende Ursache relativ zu V. Die Monotonie der Relation der beitragenden Verursachung hat daher auch Konsequenzen für direkte Verursachung: Eine direkte Ursache X für eine Variable Y relativ zu einer Variablenmenge V kann zwar relativ zu einer Erweiterung V + von V aufhören, eine *direkte* Ursache für Y zu sein. Jedoch wird X dann relativ zu V + noch immer eine *beitragende* Ursache für Y sein.[166] Kausalverhältnisse zwischen Variablen können also ihren Status der ‚Direktheit' verlieren, sie können jedoch nicht ganz verschwinden.

---

[165] Vgl. Woodward (2008a: 209). Strevens (2007) und Strevens (2008) argumentiert gegen diese Annahme der Monotonie. Wie in McCain (2015) und Statham (2017) überzeugend gezeigt wird, funktioniert sein Einwand jedoch nicht: Er beruht auf falschen Annahmen über den Begriff der Intervention. Andere Probleme für die Monotonie beitragender Verursachung ergeben sich jedoch, wie noch deutlich werden wird, aus Supervenienz-Beziehungen zwischen Variablen.

[166] Vgl. Woodward (2003: 56).

Sowohl (DU) als auch (BU) sind also auf Variablenmengen relativiert. In der Definition einer Interventionsvariable (IV) hingegen macht Woodward von kausalen Begriffen Gebrauch, die keine explizite Relativierung auf Variablenmengen enthalten. Er verlangt zum Beispiel, dass Interventionsvariablen auf X hinsichtlich Y statistisch unabhängig von allen Variablen sind, die Y *verursachen* (und nicht auf einem kausalen Pfad von X zu Y liegen) – und nicht, dass eine Interventionsvariable auf X hinsichtlich Y statistisch unabhängig von allen Variablen ist, die Y *relativ zu einer Variablenmenge V* verursachen (und nicht auf einem kausalen Pfad von X zu Y *relativ zu V* liegen). Auch der Begriff der Interventionsvariable selbst ist nicht auf Variablenmengen relativiert. Dies wirft die Frage auf, wie die nicht-relativierte Verwendung kausaler Begriffe in (IV) sich zu den relativierten Begriffen der direkten und beitragenden Verursachung verhält.

Eine mögliche Antwort hierauf ist die folgende:[167] Ausgehend von den auf Variablenmengen relativierten Begriffen der direkten und beitragenden Verursachung kann man einen de-relativierten Begriff der Verursachung definieren. Die Definition einer Interventionsvariablen bezieht sich dann auf diesen derelativierten Begriff der Verursachung.

Zur Einführung eines de-relativierten Begriffs der Verursachung bietet es sich an, auf die eine oder andere Weise über Variablenmengen zu quantifizieren. Hier eröffnen sich zwei auf den ersten Blick plausible Möglichkeiten:

> **De-Relativierung durch Allquantifikation**: Eine Variable X ist eine Ursache für eine Variable Y genau dann, wenn gilt: X ist eine beitragende Ursache für Y relativ zu jeder Variablenmenge V, die X und Y enthält.
>
> **De-Relativierung durch existentielle Quantifikation**: Eine Variable X ist eine Ursache für eine Variable Y genau dann, wenn gilt: Es gibt eine Variablenmenge V, relativ zu der X eine beitragende Ursache für Y ist.

Die De-Relativierung durch Allquantifikation ist jedoch bei genauerem Hinsehen ungeeignet. Grund hierfür ist die oben schon ausgeführte Beobachtung, dass es möglich ist, dass X keine beitragende Ursache für Y relativ zu V ist, aber relativ zu einer Erweiterung V + von V zu einer beitragenden Ursache für Y wird. Würde die De-Relativierung durch Allquantifikation gewählt, würde die Variable E im Eisenbahn-Beispiel (der Eisenbahnbau) fälschlicherweise nicht als (simpliciter) Ursache für T (Transportkosten) zählen. Denn E ist nicht beitragende Ursache für T relativ zu {E, T}, sondern erst relativ zu {E, T, A}. E ist also nicht beitragende Ursache für T relativ zu jeder Variablenmenge, die E und T

[167] Eine andere Option würde darin bestehen, auch den Begriff einer Interventionsvariablen auf Variablenmengen zu relativieren. Vgl. hierzu z. B. Weslake (2011).

enthält. Daher ist E nach Maßgabe der De-Relativierung durch Allquantifikation keine Ursache für T *simpliciter*. Allgemein würde eine De-Relativierung durch Allquantifikation dazu führen, dass nur totale Ursachen als simpliciter Ursachen gezählt werden.[168]

Woodward präferiert entsprechend eine De-Relativierung durch existentielle Quantifikation.[169] Hier ist E im Eisenbahn-Beispiel eine simpliciter Ursache für T. Denn es gibt eine Variablenmenge – in diesem Falle zum Beispiel {E, T, A} – relativ zu der E eine beitragende Ursache für T ist.[170] Dies gibt uns also den ausschlaggebenden Begriff der Verursachung im Interventionismus: Eine Variable X ist genau dann eine Ursache für eine Variable Y, wenn es eine Variablenmenge gibt, relativ zu der X eine beitragende Ursache für Y ist.

Für die Plausibilität dieser Art der De-Relativierung ist es essentiell, dass die Relation der (relativierten) beitragenden Verursachung im oben geschilderten Sinne monoton ist: Die Hinzufügung weiterer Variablen zu einer Variablenmenge darf nie dazu führen, dass Beziehungen der beitragenden Verursachung verloren gehen. Denn anderenfalls wären wir (gelegentlich) mit der folgenden Situation konfrontiert: X ist eine beitragende Ursache für Y relativ zu einer Variablenmenge V und daher ist X eine Ursache für Y simpliciter. Es gibt jedoch eine Erweiterung V + von V, relativ zu der X keine beitragende Ursache für Y ist. Dies spricht nicht dagegen, dass X eine Ursache für Y simpliciter ist. Wir wären entsprechend darauf angewiesen, zu behaupten, dass die Erweiterung V + von V die (de-relativierten) kausalen Fakten falsch wiedergibt, während nur die ‚ärmere' Variablenmenge V eine korrekte Wiedergabe der Kausalverhältnisse ermöglicht. Dies scheint jedoch äußerst unplausibel: Denn wieso sollte eine korrekte Repräsentation der kausalen Fakten zu einer falschen führen, wenn weitere Elemente der Realität zusätzlich in Betracht gezogen werden?

---

[168] Ein weiteres Problem ist, dass keine begrenzte Zahl kausaler Modelle einen Schluss darauf erlauben würde, dass X Y verursacht. Daraus, dass X eine beitragende Ursache für Y relativ zu einer Variablenmenge V ist, können wir nicht folgern, dass X eine Ursache für Y ist.

[169] Vgl. Woodward (2008b: 209).

[170] Ein scheinbarer Nachteil der De-Relativierung durch existentielle Quantifikation besteht darin, dass sie keine Begründung *der Falschheit von* Kausalaussagen auf der Grundlage einer begrenzten Zahl kausaler Modelle erlaubt. Daraus, dass X keine beitragende Ursache für Y relativ zu einer Variablenmenge V ist, können wir nicht schließen, dass X keine Ursache für Y ist. Wir müssten vielmehr zeigen, dass es keine Variablenmenge V gibt, relativ zu der X eine beitragende Ursache für Y ist.

Noch ein Kommentar zum Verhältnis zwischen Verursachung in dem gerade spezifizierten Sinne und totaler Verursachung: Der Begriff der totalen Verursachung ist von vornherein nicht auf Variablenmengen relativiert. Die Relativierung und die Forderung, andere Variablen festzuhalten, kam in erster Linie deshalb ins Spiel, weil bestimmte Ursachen, wie E im Eisenbahnbeispiel, keine *totalen* Ursachen sind. Woodward macht jedoch auch deutlich, dass alle totalen Ursachen Ursachen sind. Tatsächlich ist totale Verursachung hinreichend für die hier spezifizierte Relation der Verursachung: Wenn eine Variable X eine totale Ursache für eine Variable Y ist, dann gibt es auch eine Variablenmenge – zum Beispiel {X, Y} – relativ zu der X eine beitragende Ursache für Y ist. Um positive Kausalaussagen zu begründen, reicht es daher oft, sich auf den (deutlich leichter zu handhabenden) Begriff der totalen Verursachung zu berufen. Die Komplikationen, die durch die Relativierung auf Variablenmengen und die darauffolgende De-relativierung entstehen, können dann umgangen werden.

Soweit haben wir also eine de-relativierte Definition von Verursachung zwischen Variablen. Meine weitere Diskussion wird sich in erster Linie an dieser Definition orientieren. Es sei jedoch an dieser Stelle darauf hingewiesen, dass die Relation der Verursachung zwischen Variablen nicht gleichgesetzt werden kann mit einer Relation der aktualen Verursachung, die zwischen tatsächlich stattfindenden Ereignissen besteht. Woodward entwickelt aufbauend auf der Verursachungsrelation zwischen Variablen auch eine Definition aktualer Verursachung. Diese besteht nicht zwischen Variablen, sondern zwischen den aktualen Werten von Variablen. Die Details seiner Definition müssen uns hier nicht interessieren.[171] Zwei Dinge werden jedoch in der weiteren Diskussion noch einmal wichtig:

Erstens: Wenn eine Variable X eine Variable Y verursacht, dann werden in aller Regel zumindest manche Werte von X in manchen Situationen auch manche Werte von Y *aktual* verursachen. Die Verursachungsrelation zwischen Variablen erlaubt also durchaus – vorsichtige – Schlüsse auf aktuale Verursachung zwischen Variablenwerten.

Zweitens: Ein Wert x einer Variablen X kann nur dann eine aktuale Ursache für einen Wert y einer Variablen Y sein, wenn X eine Ursache für Y ist. Wenn X also Y *nicht* verursacht, ist der Schluss darauf, dass die Werte von X die Werte von Y nicht aktual verursachen, unproblematisch und erlaubt.

[171] Vgl. hierfür Woodward (2003: 74–86).

Ausgerüstet mit diesem komplexen System von Definitionen kausaler Begriffe werde ich nun in den kommenden Abschnitten diskutieren, wie der Interventionismus auf die Frage nach der Existenz mentaler Verursachung im nicht-reduktiven Physikalismus angewendet werden kann.

4.4.4 Der Interventionismus als Rettung für den nicht-reduktiven Physikalismus?

Der Interventionismus scheint zunächst gut zum nicht-reduktiven Physikalismus zu passen. Dieser Einschätzung verleiht auch Woodward in der folgenden Passage Ausdruck:

> Prima facie, [interventionism] seems to support the claim that mental states can be causes. We do after all seem to regularly (and successfully) intervene to change the mental states of others and perhaps our own mental states as well and these changes in turn sometimes seem to be regularly associated with changes in other mental states and behavior. Indeed, this seems to be what successful persuasion and deception are all about – in persuasion, I manipulate your beliefs and desires by providing you with information or material inducements, typically with the goal in mind that these changes will in turn lead to further changes that I desire in your mental state or behavior. On an interventionist conception of cause, this is all that is required for mental causation – nothing more metaphysically portentous is needed.[172]

Schon aus unseren alltäglichen Erfahrungen über die Interaktion mit unseren Mitmenschen wissen wir demzufolge, dass wir häufig das Verhalten anderer Personen ändern können, indem wir ihre mentalen Zustände manipulieren. Da eine solche Manipulierbarkeitsbeziehung nach Maßgabe des Interventionismus bereits hinreichend für eine Kausalbeziehung zwischen mentalen Ereignissen und Verhaltensereignissen ist, verspricht der Interventionismus somit eine einfache Erwiderung auf den Vorwurf des Epiphänomenalismus: Der nicht-reduktive Physikalismus führt nicht dazu, dass es keine mentale Verursachung gibt, weil der nicht-reduktive Physikalismus nicht dazu führt, dass wir das Verhalten anderer Personen nicht durch eine Manipulation ihrer mentalen Zustände verändern können.

Dieser auf den ersten Blick durchaus plausiblen Grundidee folgen neben Woodward auch eine Reihe von anderen Autor*innen.[173] Jedoch bleiben einige Fragen offen:

---

[172] Woodward (2008a: 231). Vgl. auch Woodward (2015: 321).

[173] Vgl. z. B. Shapiro und Sober (2007), Raatikainen (2010).

Erstens zieht die von Woodward angeführte und intuitiv einleuchtende Überlegung zur Manipulierbarkeit des Verhaltens von Personen durch eine Manipulation ihrer mentalen Zustände die Besonderheiten des nicht-reduktiven Physikalismus zunächst gar nicht in Betracht. Man könnte Woodward als Kritiker*in des nicht-reduktiven Physikalismus entsprechend durchaus darin zustimmen, dass eine solche Manipulierbarkeitsbeziehung offenbar besteht, und eben anfügen, dass diesem Sachverhalt nur Rechnung getragen werden kann, wenn der nicht-reduktive Physikalismus falsch ist. Dass die Existenz mentaler Verursachung unter Voraussetzung des Interventionismus offensichtlich ist, zeigt noch nicht, dass der Interventionismus die Existenz mentaler Verursachung *im Rahmen des nicht-reduktiven Physikalismus offensichtlich einfangen kann.*

Hiermit zusammenhängend ergeben sich zweitens aus dem nicht-reduktiven Physikalismus durchaus schwierige Interpretationsfragen für den Interventionismus. Denn der nicht-reduktive Physikalismus postuliert Beziehungen der nicht-kausalen, ontologischen Abhängigkeit zwischen niedrigerstufigen und höherstufigen Ereignissen und Eigenschaften. Der Interventionismus hingegen wurde entwickelt, ohne die Möglichkeit solcher nicht-kausalen Beziehungen im Blick zu haben. Er konzentriert sich auf kausale Beziehungen. Was aber geschieht, wenn die durch den Interventionismus bestimmten kausalen Beziehungen auf die vom nicht-reduktiven Physikalismus postulierten nicht-kausalen Beziehungen ontologischer Abhängigkeit treffen? Wie sich in den kommenden Abschnitten zeigen wird, ist es keinesfalls eine triviale Frage, ob der Interventionismus mit solchen nicht-kausalen Abhängigkeitsbeziehungen umgehen kann und zugleich zum aus Sicht des nicht-reduktiven Physikalismus gewünschten Ergebnis führt.

4.4.5 Unmögliche Kombinationen von Variablenwerten und Variablen-Supervenienz

Die nicht-kausalen ontologischen Abhängigkeitsbeziehungen, die laut dem nicht-reduktiven Physikalismus zwischen höherstufigen Ereignissen und ihren Basen bestehen, haben (in der Terminologie des Interventionismus) zur Konsequenz, dass bestimmte Kombinationen von Werten von Variablen metaphysisch unmöglich sind. Um dies zu präzisieren, diskutiere ich im Folgenden, wie man eine für den nicht-reduktiven Physikalismus typische Situation der Instantiierung einer mentalen Eigenschaft mithilfe des interventionistischen Vokabulars von Variablen und ihren Werten modellieren kann und welche Kombinationen von Variablenwerten dann als unmöglich ausgezeichnet werden müssen.

Gehen wir davon aus, dass **M** eine mentale Eigenschaft ist.[174] **M** könnte, wie im Beispiel von Susi, etwa für Susis Wunsch stehen, die Aufmerksamkeit der Dozentin auf sich zu lenken. **M** ist laut dem nicht-reduktiven Physikalismus multipel realisierbar. Gehen wir vereinfachend davon aus, dass **M** nur zwei mögliche totale[175] physische$_{eng}$ Basen hat: **BM1** und **BM2**.Im nicht-reduktiven Physikalismus necessitieren diese physischen Basen jeweils **M**: Wenn **BM1** instantiiert ist, muss **M** instantiiert sein. Entsprechendes gilt für **BM2**. Zudem gehen wir davon aus, dass **M** nicht ohne physische Basis instantiiert sein kann. Das heißt, wenn **M** instantiiert ist, muss **BM1** oder **BM2** instantiiert sein.

Um diese Situation zu modellieren, können wir nun Variablen einführen, die die Instantiierung von **M**, **BM1** und **BM2** anzeigen. Zur Repräsentation von **M** empfiehlt sich die Einführung einer binären Variable M, die in Tabelle 4.3 spezifiziert ist:

Tabelle 4.3 Variable M

Variable	Werte	Interpretation der Werte
M	m0, m1	m0 = die mentale Eigenschaft **M** ist nicht instantiiert. m1 = die mentale Eigenschaft **M** ist instantiiert.

Zur Modellierung der physischen Basen von **M** eröffnen sich zwei verschiedene Optionen. Die erste Option besteht darin, eine dreiwertige Variable BM einzuführen (siehe Tabelle 4.4).[176]

Variable BM verstößt strenggenommen gegen das in Abschnitt 4.4.3. erwähnte absolute Kriterium der Exklusivität. Der nicht-reduktive Physikalismus schließt nicht grundsätzlich aus, dass mehrere totale Basen einer mentalen Eigenschaft zugleich instantiiert sind. In diesem Falle wäre M ‚übernecessitiert' oder ontologisch überdeterminiert, ähnlich wie Wirkungen gelegentlich kausal überdeterminiert sind. Entsprechend setzt die hier verwendete Modellierung voraus,

[174] Ich verwende im Folgenden fettgedruckte Buchstaben, um Eigenschaften zu bezeichnen. Nicht fettgedruckte Buchstaben bezeichnen Variablen.

[175] Würde die Annahme, dass es sich bei **BM1** und **BM2** um totale Basen (und nicht bloß Kernbasen) handelt, aufgegeben, könnte man die relevanten Situationen durch Hinzunahme einer Variablen modellieren, die eine Reihe von günstigen oder ungünstigen Hintergrundbedingungen repräsentiert. Die unmöglichen Kombinationen von Variablenwerten würden dann immer auch diese Hintergrundvariable involvieren.

[176] Shapiro und Sober (2007) sowie Baumgartner (2009) und Baumgartner (2010) wählen beispielsweise eine Modellierung, die verschiedene mögliche Basen von M in einer einzigen, mehrwertigen Variable repräsentiert.

Tabelle 4.4 Variable BM

Variable	Werte	Interpretation der Werte
BM	bm0, bm1, bm2	bm0 = Es ist keine totale Basis von **M** instantiiert. bm1 = Die physische Eigenschaft **BM1**, die eine totale Basis von **M** ist, ist instantiiert. bm2 = Die physische Eigenschaft **BM2**, die eine totale Basis von **M** ist, ist instantiiert.

dass diese Möglichkeit der ontologischen Überdetermination im gegebenen Kontext vernachlässigt werden kann. Anderenfalls müsste man die Möglichkeit in Betracht ziehen, dass BM sowohl den Wert bm1 als auch den Wert bm2 annimmt. Dies sollte jedoch vermieden werden.

Nimmt man die Möglichkeit ontologischer Überdetermination also ernst, ist es angeraten, für jede mögliche Base von **M** eine eigene binäre Variable einzuführen. Eine zweite, alternative Modellierung der Situation beinhaltet also die folgenden beiden, binären Variablen BM1 und BM2 (siehe Tabelle 4.5).[177]

Tabelle 4.5 Variablenmenge {BM1, BM2}

Variable	Werte	Interpretation der Werte
BM1	bm10, bm11	bm10 = die physische Eigenschaft **BM1**, die eine totale Basis von **M** ist, ist nicht instantiiert. bm11 = die physische Eigenschaft **BM1**, die eine totale Basis von **M** ist, ist instantiiert.
BM2	bm20, bm21	bm20 = die physische Eigenschaft **BM2**, die eine totale Basis von **M** ist, ist nicht instantiiert. bm21 = die physische Eigenschaft **BM2**, die eine totale Basis von **M** ist, ist instantiiert.

Bei beiden Optionen ergeben sich bestimmte Kombinationen zwischen Werten der eingeführten Variablen, die im nicht-reduktiven Physikalismus als metaphysisch unmöglich gelten müssen. So sind, bezogen auf die erste Option, die folgenden Kombinationen von Werten von M und BM metaphysisch unmöglich:

(Kom-1) BM=bm0 und M=m1

(Kom-2) BM=bm1 und M=m0

[177] Eine solche Repräsentation, in der verschiedene mögliche Basen von M durch verschiedene, binäre Variablen repräsentiert sind, findet sich etwa in Pernu (2013).

(Kom-3) BM=bm2 und M=m0

Kombination (Kom-1) entspräche einer Situation, in der **M** ohne eine Basis instantiiert ist. Die Kombinationen (Kom-2) und (Kom-3) entsprächen jeweils einer Situation, in der eine totale Basis von M instantiiert ist, ohne dass M instantiiert ist. Wie oben geschildert, werden genau solche Situationen durch den nicht-reduktiven Physikalismus als metaphysisch unmöglich ausgeschlossen.

Bei der zweiten Option zur Modellierung von **M** und ihren Basen ist die Sache nur leicht komplizierter. Betrachten wir zunächst die Beziehung zwischen M und BM1. Hier stellt sich nur die folgende Kombination als metaphysisch unmöglich heraus:

(Kom-4) BM1=bm11 und M=m0[178]

Denn (Kom-4) würde einer Situation entsprechen, in der eine totale Basis von **M** ohne **M** instantiiert ist. Jede andere Kombination von Werten von M und BM1 ist hingegen metaphysisch möglich. Insbesondere ist es möglich, dass M den Wert m1 annimmt und BM1 den Wert bm10 annimmt. **M** wäre dann instantiiert, ohne dass die spezifische Base **BM1** von **M** instantiiert ist. Da **M** multipel realisierbar ist, ist diese Situation möglich.

Da **M** aber, wie wir voraussetzen, notwendigerweise eine Basis hat und **BM2** die einzige mögliche alternative Basis ist, muss dann BM2 den Wert bm21 annehmen. Das heißt: Wenn wir alle drei Variablen M, BM1 und BM2 betrachten, ergeben sich weitere unmögliche Kombinationen von Werten:

(Kom-5) BM1=bm10, BM2=bm20 und M=m1

(Kom-6) BM1=bm11, BM2=bm20 und M=m0

(Kom-7) BM1=bm10, BM2=bm21 und M=m0

(Kom-8) BM1=bm11, BM2=bm21 und M=m0

Hier entspricht (Kom-5) einer Situation, in der **M** ohne totale Basis instantiiert ist, während (Kom-6), (Kom-7) und (Kom-8) Situationen entsprechen, in denen (mindestens) eine totale Basis von **M** ohne **M** instantiiert ist. Wieder müssen die entsprechenden Kombinationen von Werten nach den Vorgaben des

---

[178] Aus demselben Grund ist auch die folgende Kombination von Werten von M und BM2 metaphysisch unmöglich: BM2 = bm21 und M = m0. Auch alles weitere, was ich hier über die Relation zwischen M und BM1 sage, überträgt sich eins-zu-eins auf die Relation zwischen M und BM2.

nicht-reduktiven Physikalismus also als metaphysisch unmöglich ausgeschlossen werden.

Wenn Variablen so aufeinander bezogen sind, wie M auf BM bzw. auf BM1 und BM2, können wir von einer *Supervenienz-Relation zwischen Variablen* sprechen. Diese kann wie folgt eingeführt werden:

> **Variablen-Supervenienz (VS)**: Eine Variable X superveniert genau dann auf einer Variablenmenge {Y1, …, Yn}, wenn gilt: Für jeden Wert x von X gibt es eine Kombination von Werten k von Y1, …, Yn, so dass notwendigerweise gilt: Wenn Y1, …, Yn die Kombination von Werten k annimmt, dann nimmt X den Wert x an.[179]

M superveniert auf BM[180], weil es für jeden Wert von M einen Wert von BM gibt, der ihn necessitiert: Der Wert m0 wird durch den Wert bm0 necessitiert: Wenn BM den Wert bm0 annimmt, dann muss M den Wert m0 annehmen. Der Wert m1 wird sowohl von bm1 als auch von bm2 necessitiert: Wenn BM den Wert bm1 annimmt, muss M den Wert m1 annehmen, und wenn BM den Wert bm2 annimmt, muss M ebenfalls den Wert m1 annehmen. Entsprechend gibt es für jeden Wert von M (mindestens) einen Wert von BM, der ihn necessitiert. Anders ausgedrückt: Es ist unmöglich, M zu ändern, ohne BM zu ändern. Wenn M sich von m0 zu m1 ändert, muss sich auch BM von bm0 auf bm1 oder bm2 ändern. Und wenn M sich von m1 zu m0 ändert, muss sich auch BM von bm1 oder bm2 zu bm0 ändern.

M superveniert zudem auf der Variablenmenge {BM1, BM2}. Für jeden Wert von M gibt es eine Kombination von Werten von BM1 und BM2, der ihn necessitiert. Für den Wert m0 ist dies die Kombination (bm10 und bm20). Für den Wert m1 sind dies die verbleibenden drei Kombinationen (bm11 und bm20), (bm10 und bm21) sowie (bm11 und bm21). Jedoch superveniert M weder auf BM1 noch auf BM2 alleine: Denn es gibt keinen Wert von BM1, der m0 necessitiert. Der Wert bm10 necessitiert nicht m0, weil M = m1 sein kann, während BM1 = bm10 ist, (insofern BM2 = bm21 ist). Eine Supervenienz-Beziehung unter Voraussetzung einer binären Repräsentation der Basen von **M** muss entsprechend sämtliche Variablen für die möglichen Basen von **M** involvieren.

Bis hierher haben wir eine Übertragung der zentralen modalen Annahmen des nicht-reduktiven Physikalismus auf die Terminologie des Interventionismus. Aus dem nicht-reduktiven Physikalismus folgt, dass bestimmte Kombinationen

---

[179] Vgl. Hoffmann-Kolss (2021).

[180] Strenggenommen müsste man sagen: ‚M superveniert auf der Menge {BM}'. Die ‚Kombination' k besteht im vorliegenden Anwendungsfall jeweils in einem einzigen Wert.

von Werten von Variablen metaphysisch unmöglich sind und dass bestimmte Variablen auf anderen Variablen supervenieren.

Der Unmöglichkeit von Kombinationen von Variablenwerten zugrunde liegt dabei eine *nicht-kausale* ontologische Abhängigkeitsbeziehung. Es ist daher mit dem nicht-reduktiven Physikalismus zugleich verbunden, dass *keine* kausalen Beziehungen zwischen den Variablen M und BM, bzw. zwischen den Variablen M und BM1 und BM2 bestehen: M ist keine Ursache für BM, BM1 und BM2 und BM, BM1 und BM2 sind keine Ursachen für M. M superveniert *nicht-kausal* auf BM bzw. {BM1, BM2}. Der nicht-reduktive Physikalismus kommt also mit einer Einschränkung für die Möglichkeit von Kombinationen von Variablenwerten und zugleich mit einem Verbot, diese Einschränkung auf kausale Beziehungen zurückzuführen. Diese Vorgaben bleiben nicht ohne Konsequenzen für die Frage, in welche kausalen Beziehungen die entsprechenden Variablen laut dem Interventionismus eintreten können.

Im folgenden Abschnitt gehe ich zunächst auf ein unmittelbares Problem ein, das sich im Zusammenhang mit der Anwendung von Woodwards Definitionen der direkten und beitragenden Verursachung auf Variablen, die in Supervenienzbeziehungen stehen, ergibt. Ich beschränke mich hierbei auf eine Diskussion der Variablen M und BM und wähle also die Modellierung der ausschlaggebenden Situation über eine dreiwertige Variable, die die Anwesenheit der verschiedenen Basen von **M** anzeigt.

4.4.6 Direkte Verursachung, beitragende Verursachung und Variablen-Supervenienz

Wenn wir die Definitionen einer direkten oder beitragenden Ursache auf Variablenmengen anwenden, die Variablen enthalten, die in einer Supervenienzbeziehung stehen, ergeben sich seltsame und schwer zu akzeptierende Konsequenzen. Dies möchte ich im Folgenden ausführen.

Zuerst gehe ich dafür knapp auf die Annahme des nicht-reduktiven Physikalismus ein, dass mentale Eigenschaften oder Ereignisse und ihre physischen Basen nicht kausal verknüpft sind. Eine Anwendung der Definition der direkten und beitragenden Verursachung auf Variablenmengen, die in Supervenienzbeziehungen stehen, hat auf den ersten Blick die gegenteilige Konsequenz: Variablen, die in Supervenienzbeziehungen stehen, verursachen einander.

Betrachten wir dafür die Variablenmenge {M, BM}. Auf den ersten Blick gibt es nun Interventionen[181] auf M, die BM verändern. Tatsächlich ändert sogar *jede* Intervention auf M mit metaphysischer Notwendigkeit auch BM. Folgt man also der Definition der direkten oder der beitragenden Verursachung, ist M eine direkte und beitragende Ursache für BM relativ zu {M, BM}. Also gibt es eine Variablenmenge, relativ zu der M direkte oder beitragende Ursache für BM ist. Also ist M eine Ursache für BM.

Zudem gibt es auch Interventionen auf BM, die M verändern. Hier ändert zwar nicht jede Intervention auf BM auch M: Die Interventionen auf BM, die zwischen den Werten bm1 und bm2 wechseln, ziehen keine Änderung von M nach sich. Jedoch führt jede andere Intervention auf BM (z. B, eine solche, die BM von bm0 auf bm1 setzt) mit metaphysischer Notwendigkeit zu einer Änderung von M. Demzufolge gibt es eine Intervention auf BM, die M ändert. Also ist BM eine direkte und beitragende Ursache für M relativ zu {M, BM}. Also ist auch BM eine Ursache für M.

Aus Sicht des nicht-reduktiven Physikalismus ist dieses Ergebnis natürlich unbefriedigend. Wie oben geschildert geht mit dem nicht-reduktiven Physikalismus die Vorgabe einher, die Beziehung zwischen M und BM nicht-kausal zu deuten. Es scheint jedoch, als habe der Interventionismus bis hierher nicht die nötigen Mittel, um die nicht-kausalen Beziehungen ontologischer Abhängigkeit, die laut dem nicht-reduktiven Physikalismus zwischen höherstufigen Eigenschaften und ihren physischen Basen bestehen, von gewöhnlichen kausalen Beziehungen zu unterscheiden.[182]

Ich komme auf dieses Problem gegen Ende dieses Abschnitts noch einmal zurück. Vorher möchte ich das Problem des Interventionismus mit Variablen, die in einer nicht-kausalen Supervenienz-Beziehung stehen, jedoch noch etwas verschärfen. Dafür gehe ich vorerst davon aus, dass – trotz der gerade ausgeführten Überlegung – keine kausale Beziehung zwischen M und BM besteht. Das Problem verschärft sich, wenn man die Frage stellt, in welchen kausalen Beziehungen M und BM zu anderen Variablen stehen können, wenn man davon ausgeht, dass sie in einer *nicht-kausalen* Supervenienzbeziehung zueinander stehen.

[181] Ich ignoriere in diesem Abschnitt zunächst die Komplikationen, die sich im Zusammenhang mit der Annahme ergeben, dass es überhaupt Interventionen auf M hinsichtlich W gibt. Diese behandle ich unten im Zusammenhang mit dem interventionistischen Exklusionsprinzip aber noch ausführlich. Vorwegnehmend können die Bezüge auf Interventionen in diesem Abschnitt im Sinne von (IV*) – also des modifizierten Interventionsbegriffs, der Ausnahmeklauseln für Supervenienzbasen enthält – verstanden werden.

[182] Vgl. für eine Diskussion dieses Problems auch Hoffmann-Kolss (Manuskript).

Betrachten wir hierfür zunächst die Variablenmenge {M, W}. M ist wie oben bestimmt und zeigt an, ob die mentale Eigenschaft **M** instantiiert ist. W zeigt an, ob eine bestimmte behaviorale Eigenschaft **W** instantiiert ist. **W** kann dabei zum Beispiel so verstanden werden, dass eine Instantiierung von **W** bedeutet, dass Susi ihren Arm hebt. Die Variable W ist also wie in Tabelle 4.6 bestimmt.

Tabelle 4.6 Variable W

Variable	Werte	Interpretation der Werte
W	w0, w1	w0 = die behaviorale Eigenschaft **W** ist nicht instantiiert. w1 = die behaviorale Eigenschaft **W** ist instantiiert.

Gehen wir nun davon aus, dass eine reine Betrachtung von M und W nahelegt, dass M eine Ursache für W ist. Tatsächlich sieht alles danach aus, dass M eine direkte Ursache für W relativ zu {M, W} ist und die Beziehung zwischen M und W durch die einfache Strukturgleichung ‚$W: = F_W(M)$‘ beschrieben werden kann (siehe Tabelle 4.7). Das heißt, Interventionen[183] auf M, die M auf m0 setzen, führen dazu, dass W den Wert w0 annimmt und Interventionen, die M auf m1 setzen, führen dazu, dass W den Wert w1 annimmt.

Tabelle 4.7 Funktionstabelle für Gleichung $W := F_W(M)$

M	$F_W(M)$
m0	w0
m1	w1

Gegeben Woodwards Definition der direkten Verursachung wäre M dann eine direkte Ursache für W relativ zu {M, W}: Eine Intervention auf M verändert W, wenn alle anderen Variablen in {M, W} auf einem Wert festgehalten werden.

Betrachten wir als nächstes die Variablenmenge {BM, W}, wobei BM wie in Tabelle 4.4 definiert anzeigt, welche physische Basis von M instantiiert ist. Auch hier sieht alles danach aus, dass BM eine direkte Ursache für W relativ zu {BM, W} ist und die Beziehung zwischen BM und W durch die Strukturgleichung

[183] Ich ignoriere in diesem Abschnitt zunächst die Komplikationen, die sich im Zusammenhang mit der Annahme ergeben, dass es überhaupt Interventionen auf M hinsichtlich W gibt. Diese behandle ich unten im Zusammenhang mit dem interventionistischen Exklusionsprinzip aber noch ausführlich. Vorwegnehmend können die Bezüge auf Interventionen in diesem Abschnitt im Sinne von (IV*) – also des modifizierten Interventionsbegriffs, der Ausnahmeklauseln für Supervenienzbasen enthält – verstanden werden.

‚W: = F_w(BM)‘ beschrieben werden kann (siehe Tabelle 4.8): Interventionen, die BM auf bm0 setzen, haben zur Folge, dass W den Wert w0 annimmt. Interventionen, die BM auf bm1 oder bm2 setzen, haben zur Folge, dass W den Wert w1 annimmt. Man beachte, dass hier nicht jede Intervention auf BM zu einer Änderung von W führt. Interventionen, die zwischen den Werten bm1 und bm2 wechseln, führen zu keiner Änderung von W. Dies ist jedoch wie oben beschrieben mit Woodwards Definition einer direkten Ursache kompatibel: Schließlich gibt es eine Intervention auf BM, die zu einer Änderung von W führt, wenn alle anderen Variablen in {BM, W} auf einem Wert festgehalten werden.

Tabelle 4.8 Funktionstabelle für Gleichung W := Fw(BM)

BM	F_w(BM)
bm0	w0
bm1	w1
bm2	w1

Die so bestimmten (scheinbar) kausalen Zusammenhänge zwischen BM und W bzw. M und W sind zunächst kompatibel mit der Annahme, dass M auf BM superveniert: Da (i) M notwendigerweise genau dann den Wert m0 hat, wenn BM den Wert m0 hat und W laut den gerade spezifizierten Strukturgleichungen dann den Wert w0 hat und (ii) M notwendigerweise genau dann den Wert m1 hat wenn BM den Wert bm1 oder den Wert bm2 hat und W laut den gerade spezifizierten Strukturgleichungen dann den Wert w1 hat, ist hier keine Inkonsistenz zu finden. Zugleich ist jedoch zu beachten, dass andere Strukturgleichungen für BM und W bzw. M und W nicht mit der Supervenienz von M auf BM vereinbar wären. Insbesondere kann, wenn der Zusammenhang zwischen M und W durch ‚W: = F_w(M)‘ bestimmt ist und M auf BM superveniert, keiner der Werte bm1 und bm2 von BM dazu führen, dass W den Wert w0 annimmt. Gegeben die Strukturgleichung ‚W: = F_w(M)‘ für {M, W} können Interventionen, die zwischen den physischen Basen von **M** wechseln, nicht zu einer Änderung von W führen. In anderem Vokabular ausgedrückt: Die Strukturgleichung ‚W: = F_w(M)‘ legt fest, dass die Beziehung zwischen M und W *realisierungsinsensitiv* ist.[184]

Nun stellt sich jedoch die Frage, was passiert, wenn wir die bisher betrachteten Modelle zusammenführen. Was passiert also, wenn wir statt den Variablenmengen {M, W} und {BM, W} die Erweiterung {M, BM, W} betrachten?

Unter der in Abschnitt 4.4.3. schon behandelten Annahme der Monotonie sollte zu erwarten sein, dass die kausalen Beziehungen zwischen M und W bzw.

[184] Vgl. auch Woodward (2008a: 241).

zwischen BM und W, die relativ zur Variablenmenge {M, W} bzw. {BM, W} aufzufinden sind, relativ zu der Erweiterung {M, BM, W} nicht verloren gehen. Gehen wir also davon aus, dass M relativ zu {M, W} eine direkte Ursache für W ist, muss M auch eine direkte (oder beitragende) Ursache für W relativ zu {M, BM, W} sein. Entsprechendes gilt für BM. Aus dieser Perspektive ist also zu erwarten, dass sowohl M als auch BM direkte (oder beitragende) Ursachen für W relativ zu {M, BM, W} sind.

Aus einer naiven Anwendung der Definition der direkten Verursachung (DU) bzw. der beitragenden Verursachung (BU) auf die Variablenmenge {M, BM, W} ergibt sich jedoch recht unmittelbar, dass *weder* M *noch* BM direkte oder beitragende Ursachen für W relativ zu {M, BM, W} sind. Ausschlaggebend hierfür ist die sowohl in (DU) als auch (BU) enthaltene Forderung, dass es eine mögliche Intervention auf die Ursache gibt, die die Wirkung verändert, *während alle anderen Variablen in der Variablenmenge (die nicht auf dem kausalen Pfad von der Ursache zur Wirkung liegen) auf einem Wert festgehalten werden.* Gegeben die Supervenienzbeziehung zwischen M und BM ist es bei Interventionen auf M bzw. BM nicht (immer) möglich, die jeweils andere Variable festzuhalten. Wenn ein solches Festhalten anderer Variablen aber nicht möglich ist, sind (DU) und (BU) nicht erfüllt.[185]

Zunächst zur Frage, ob M eine direkte oder beitragende Ursache für W relativ zu {M, BM, W} ist. Hier liegt die Sache recht einfach: Jede Änderung von M geht aufgrund der Supervenienzbeziehung notwendig mit einer Änderung von BM einher. Man kann BM daher nicht auf einem Wert festhalten, während man eine Intervention auf M durchführt. M ist deshalb keine *direkte* Ursache für W relativ zu {M, BM, W}. Auch liegt BM nicht auf einem kausalen Pfad von M zu W – die Supervenienz-Beziehung soll schließlich nicht-kausal aufgefasst werden, so dass BM keine Ursache oder Wirkung von M ist. Entsprechend verlangt auch (BU), dass BM auf einem Wert festgehalten werden kann, während M manipuliert wird. Da dies nicht möglich ist, ist M auch keine *beitragende* Ursache für W relativ zu {M, BM, W}. Die (vermeintliche?) kausale Beziehung zwischen M und W, die mit Blick auf die Variablenmenge {M, W} festgestellt wurde, geht also verloren, wenn man auch die Variable BM ins Boot holt.

Hinsichtlich der Frage, ob BM eine direkte oder beitragende Ursache für W relativ zu {M, BM, W} ist, ist die Argumentation nur unwesentlich komplizierter: Gegeben, dass W systematisch mit M korreliert ist und immer dann den Wert w0 bzw. w1 annimmt, wenn M den Wert m0 bzw. m1 annimmt, kann auch BM keine direkte Ursache für W relativ zu {M, BM, W} sein. Zwar gibt es Interventionen

[185] Vgl. Baumgartner (2009: 165–167) für Diskussion.

auf BM, die es erlauben, M festzuhalten. So kann man BM von bm1 auf bm2 setzen und andersherum, während M auf dem Wert m1 ‚festgehalten‘[186] wird. Jedoch sind gerade dies die Interventionen, die W nicht verändern. Wenn man auf BM hingegen so interveniert, dass sich W verändert, kann man M nicht festhalten: Setzt man BM beispielsweise von bm1 oder bm2 auf bm0 (so dass sich W von w1 auf w0 verändert), muss sich mit Notwendigkeit auch M von m1 auf m0 verändern. Dementsprechend ist jede Intervention auf BM entweder so beschaffen, dass sie W nicht verändert, oder so beschaffen, dass es nicht möglich ist, M festzuhalten. Daher ist auch BM keine Ursache für W relativ zu {M, BM, W}.

Gegeben die spezifizierten Zusammenhänge zwischen BM und W bzw. M und W und die nicht-kausale Supervenienz-Beziehung zwischen M und BM zählen also weder M noch BM als Ursache für W relativ zu {M, BM, W}.

Dieses Ergebnis kann verallgemeinert werden: Erstens wird eine Variable X niemals als Ursache für irgendeine Variable Y relativ zu einer Variablenmenge zählen, die eine Variable Z enthält, so dass X auf Z superveniert. Denn man kann allgemein Z nicht auf einem Wert festhalten, während man X verändert, wenn X auf Z superveniert. Zweitens wird eine Variable Z nur dann eine Variable Y relativ zu einer Variablenmenge verursachen, die eine Variable X enthält, die auf Z superveniert, wenn es Änderungen von Z gibt, die (i) X nicht verändern und daher mit einem Festhalten von X vereinbar sind und die (ii) Y verändern. Denn anderenfalls sind die einzigen Veränderungen von Z, die durchgeführt werden können, während X festgehalten wird, solche, die Y nicht verändern.

Einige Kommentare und Erläuterungen zu diesem Ergebnis sind angebracht:

Erstens: Es gibt durchaus Situationen, in denen eine subveniente Variable Ursachen relativ zu einer Variablenmenge hat, die auch eine zugehörige supervenierende Variable enthält. Wenn die Interventionen auf BM, die zwischen bm1 und bm2 wechseln, zu Änderungen von W führen, muss BM als Ursache für W relativ zu {M, BM, W} zählen. Hieraus ergibt sich also eine Asymmetrie zwischen supervenierenden und subvenienten Variablen: Für supervenierende Variablen ist es *allgemein* ausgeschlossen, dass sie Wirkungen relativ zu Variablenmengen haben, die auch eine zugehörige subveniente Variable enthalten. Für subveniente Variablen ist dies hingegen nur unter der Bedingung ausgeschlossen, dass die anderen Variablen in dem Modell nicht auf bestimmte Änderungen der subvenienten Variable reagieren. Man kann jedoch davon ausgehen, dass diese

[186] Wobei das ‚Festhalten‘ hier schlichtweg in der Beobachtung des Unvermeidbaren besteht. Es wäre nicht möglich, M *nicht* auf diesem Wert ‚festzuhalten‘.

Bedingung bei den paradigmatischen Fällen (vermeintlicher) mentaler Verursachung erfüllt ist: Wenn **W** im obigen Beispiel für eine typische behaviorale Eigenschaft wie Susis Armbewegung steht, kann man davon ausgehen, dass W den Wert w1 annimmt (**W** instantiiert ist), unabhängig davon, wie **M** realisiert ist. Solange die Beziehung zwischen M und W in diesem Sinne *realisierungsinsensitiv* ist, kann BM also nicht als Ursache für W relativ zu {M, BM, W} zählen.

Zweitens: Die ausgeführte Argumentation hat eine enge Verwandtschaft zu *einem* Strang[187] in Michael Baumgartners Überlegungen zum interventionistischen Exklusionsargument, die weiter unten noch ausführlich diskutiert werden. Jedoch sollte auf zumindest zwei wichtige Unterschiede hingewiesen werden: Zum ersten stecken – anders als bei Baumgartner – in der ausgeführten Argumentation keine vorherigen Annahmen über Kausalbeziehungen zwischen BM und W relativ zu {M, BM, W}. Baumgartner setzt voraus, dass BM eine Ursache für W relativ zu {M, BM, W} ist und folgert hieraus auf der Grundlage u. a. der Unmöglichkeit des Festhaltens von BM bei Interventionen auf M, dass M keine Ursache für W relativ zu {M, BM, W} ist. Wie wir gesehen haben, ist die Annahme, dass BM eine Ursache für W relativ zu {M, BM, W} ist, für dieses Ergebnis jedoch gar nicht nötig. Es reicht bereits, dass sich BM überhaupt in der fraglichen Variablenmenge befindet.[188] Zum zweiten problematisiert die vorliegende Argumentation auch diese Annahme, dass BM eine Ursache für W relativ zu {M, BM, W} ist. Baumgartner hält diese Annahme für unproblematisch und meint, dass nicht-reduktive Physikalist*innen auf sie festgelegt sind. Diese Annahme kann jedoch nur dann konsistent vertreten werden, wenn die korrelative Beziehung zwischen M und W nicht realisierungsinsensitiv ist.

Drittens: Wenn wir davon ausgehen, dass M eine Ursache für W relativ zu {M, W} ist oder BM eine Ursache für W relativ zu {BM, W} ist, dann geht das vorliegende Ergebnis mit einer Verletzung der Monotonie der beitragenden Verursachung einher. M ist direkte (und damit beitragende) Ursache für W relativ zu {M, W}, jedoch ist M nicht direkte oder beitragende Ursache für W relativ zu der erweiterten Variablenmenge {M, BM, W}. Entsprechendes gilt für BM. Man kann an der Annahme der Monotonie also auf den ersten Blick nur dadurch festhalten, dass man M auch relativ zu {M, W} nicht als Ursache für W klassifiziert

[187] Baumgartner argumentiert für das interventionistische Exklusionsargument sowohl über die Unmöglichkeit von Interventionen als auch über die Unmöglichkeit des Festhaltens anderer Variablen. Hier geht es zunächst nur um die Unmöglichkeit des Festhaltens anderer Variablen. Vgl. Baumgartner (2009) und Baumgartner (2010).

[188] Ich ziehe gleich noch die Möglichkeit in Betracht, die Definitionen der direkten und beitragenden Verursachung so abzuändern, dass die Annahme nötig wird.

und BM auch relativ zu {BM, W} nicht als Ursache für W klassifiziert. Wenn wir also an der Annahme der Monotonie festhalten, können wir das Ergebnis auf dieser Grundlage ausweiten: Es gibt dann keine Variablenmengen, relativ zu denen M oder BM Ursache für W ist.[189]

Viertens: Nur unter Voraussetzung einer solchen Ausweitung ist es möglich, aus dem vorliegenden Ergebnis de-relativierte negative Kausalaussagen abzuleiten. Gegeben die unter 4.4.3. eingeführte De-Relativierung durch existentielle Quantifikation können wir daraus, dass M keine Ursache für W relativ zu {M, BM, W} ist, nicht ableiten, dass M keine Ursache für W simpliciter ist. Ganz im Gegenteil können wir unter der Annahme, dass M eine Ursache für W relativ zu {M, W} ist, sogar ableiten, dass M eine Ursache für W simpliciter ist. Wenn wir die Annahme der Monotonie also aufgeben, hat das vorliegende Ergebnis keine Konsequenzen für de-relativierte Verursachung. Stattdessen würde die De-Relativierung durch existentielle Quantifikation dann nahelegen, dass die Variablenmenge {M, BM, W} die tatsächlichen kausalen Fakten fehlrepräsentiert, während die Variablenmengen {M, W} und {BM, W} korrekte Repräsentationen sind.

Fünftens: Die vorliegende Argumentation hat keinerlei Konsequenzen für den (von vornherein nicht relativierten) Begriff der *totalen Verursachung*: Um festzustellen, ob M eine *totale Ursache* für W ist, reicht es, eine isolierte Intervention auf M durchzuführen, ohne BM festzuhalten. Da eine solche Intervention W ändert, ist M also totale Ursache für W.[190] Gegeben, dass totale Verursachung simpliciter-Verursachung impliziert, ist M also eine Ursache für W. Dies scheint ein starkes Argument dafür zu sein, dass die Variablenmenge {M, BM, W} zu einer Fehlrepräsentation der kausalen Fakten führt und das Problem entsprechend bei der Monotonie-Annahme liegt. Halten wir nämlich an der Monotonie-Annahme fest (und gehen also davon aus, dass M keine direkte oder beitragende Ursache für W relativ zu {M, W} ist), müssen wir schließen, dass totale Verursachung nicht hinreichend für simpliciter-Verursachung ist. Hierfür gibt es jedoch zumindest keinen unabhängigen Grund.

Sechstens: Die vorliegende Argumentation beruht auf der im Wortlaut der Definitionen der direkten und beitragenden Verursachung enthaltenen Forderung, dass *alle* Variablen in der relevanten Variablenmenge, die nicht auf dem kausalen Pfad von M zu W liegen, auf einem Wert festgehalten werden müssen. Einige

[189] Man beachte, dass die Ergebnisse betreffend {M, BM, W} auch alle Erweiterungen von {M, BM, W} betreffen. Es ist daher nicht möglich, dass die Verursachungsbeziehungen bei reichhaltigeren Variablenmengen wieder auftauchen.

[190] Vgl. auch Polger, Shapiro und Stern (2018: 6).

Bemerkungen von Woodward machen jedoch deutlich, dass diese Forderung eigentlich unnötig stark ist:

> [A]s long as we hold fixed some variable along each of the alternative causal routes from X to Y besides the direct route, it does not matter what else we hold fixed. In particular, if we hold fixed *all* other variables in the variable set **V** that describes the system of interest and find that there is a possible manipulation of X that changes Y, it follows that X is a direct cause of Y with respect to that variable set.[191]

> [I]t is not really necessary to fix all off-path variables at all possible combinations of values. Instead, we may proceed as follows. We first draw a causal graph that represents all of the direct causal relationships between X, Y and all other variables in V. […] We then check all of the routes or directed paths from X to Y. For each route Pi from X to Y we freeze at least one intermediate variable at each of its various possible values along all other routes from X to Y containing intermediate variables. We also fix at each of its possible values all other direct causes of Y that are not on any direct path from X to Y. If for some combination of values of these off-path variables, some intervention on X will change the value of Y, then X is a contributing cause of Y. Alternatively, we check to see whether there is some possible combination of values of the direct causes of Y that are not on the path Pi such that with these values fixed, some intervention on X will change the value of Y.[192]

Woodward führt hier zunächst aus, dass die Forderung nach einem Festhalten *aller* Variablen in V keinen *Schaden* für seine Definition der direkten und beitragenden Verursachung anrichtet. Die gewünschten Ergebnisse im Zusammenhang mit der Unterscheidung zwischen totaler und direkter bzw. beitragender Verursachung – also in erster Linie die Klassifizierung von E im Eisenbahnbeispiel als direkte Ursache von T relativ zu {E, A, T} (siehe Abschnitt 4.4.3.) – lassen sich jedoch schon auf Grundlage einer anspruchsloseren Forderung etablieren. Insbesondere ist es für dieses Ergebnis hinreichend, dass diejenigen Variablen in {E, A, T}, die nicht auf dem direkten Pfad von E zu T liegen *und die T direkt verursachen* auf einem Wert festgehalten werden. Eine anspruchslosere Forderung an direkte und beitragende Verursachung, die ebenfalls zu den gewünschten Ergebnissen führt, wäre also die folgende:

> **Direkte Ursache – Qualifiziert (DU-Q):** Eine Variable *X* ist eine direkte Ursache einer Variablen *Y* relativ zu einer Variablenmenge V genau dann, wenn gilt: Es gibt eine mögliche Intervention auf *X* hinsichtlich *Y*, die *Y* (oder die Wahrscheinlichkeitsverteilung für Y) verändert, wenn alle anderen Variablen in V außer X und Y,

[191] Woodward (2003: 55).

[192] Woodward (2003: 59–60).

die relativ zu V direkte Ursachen von Y sind, durch Interventionen auf einem Wert festgesetzt werden.

Beitragende Ursache – Qualifiziert (BU-Q): Eine Variable X ist eine beitragende Ursache einer Variablen Y relativ zu einer Variablenmenge V genau dann, wenn gilt: (i) Es gibt einen kausalen Pfad P, d.h. eine Kette aus direkten Ursachen, der von X zu Y führt und (ii) es gibt eine Intervention auf X, die Y (oder die Wahrscheinlichkeitsverteilung für Y) verändert, wenn alle Variablen in V, die nicht auf dem kausalen Pfad P von X zu Y liegen *und die Y relativ zu V direkt verursachen*, auf einem Wert festgehalten werden.

Ein scheinbarer Nachteil dieser Definitionen ist, dass sie die Zirkularität des Interventionismus noch einmal offensichtlich werden lassen: Die Definition der direkten Verursachung nach (DU-Q) enthält selbst den Begriff der direkten Verursachung. Zugleich wäre (DU-Q) nach Woodwards erklärtem Maßstab scheinbar dennoch nicht auf problematische Weise zirkulär: Zur Anwendung von (DU-Q) benötigen wir Informationen darüber, welche anderen Variablen in V außer X direkte Ursachen für Y sind. Jedoch benötigen wir nicht die Information, ob X direkte Ursache für Y ist.[193]

Welche Konsequenzen erwachsen aus (DU-Q) und (BU-Q) für die vorliegende Argumentation betreffend die Variablenmenge {M, BM, W}?

In erster Linie ergibt sich aus diesen Definitionen, dass wir kausale Annahmen voraussetzen müssen, um zu etablieren, dass andere (relativierte) kausale Aussagen falsch sind. So lässt sich unter Voraussetzung von (DU-Q) und (BU-Q) nicht mehr direkt zeigen, dass M keine Ursache für W relativ zu {M, BM, W} ist. Stattdessen lässt sich nur noch zeigen, dass M keine Ursache für W relativ zu {M, BM, W} ist, *wenn BM eine Ursache für W relativ zu {M, BM, W} ist*. Denn das Festhalten von BM ist nur dann gefordert, wenn BM eine Ursache für W ist. Auch lässt sich nicht mehr direkt zeigen, dass BM keine Ursache für W relativ zu {M, BM, W} ist, wenn die Beziehung zwischen M und W realisierungsinsensitiv ist. Stattdessen lässt sich nur noch zeigen, dass BM keine Ursache für W relativ zu {M, BM, W} ist, wenn die Beziehung zwischen M und W realisierungsinsensitiv ist *und M eine Ursache für W relativ zu {M, BM, W} ist*. Denn das Festhalten von M ist nur dann gefordert, wenn M eine Ursache für W ist.

Eine wichtige Konsequenz hiervon ist, dass sich auf der Grundlage von (DU-Q) und (BU-Q) eine erste Version eines *interventionistischen Exklusionsprinzips* etablieren lässt. Dieses Prinzip besagt grob gesprochen, dass Variablen X und

[193] Vgl. Woodward (2003: 104–107).

Y, die in einer Supervenienzbeziehung zueinander stehen, relativ zu Variablenmengen, die X und Y enthalten, niemals gemeinsame Wirkungen haben. Etwas präziser kann es wie folgt formuliert werden:

> **Relativiertes interventionistisches Exklusionsprinzip**: Wenn eine Variable X nicht-kausal auf einer Variable Y superveniert und Y eine direkte oder beitragende Ursache für Z relativ zu einer Variablenmenge V ist, die X und Y enthält, dann gilt: X ist keine direkte oder beitragende Ursache für Z relativ zu V.

Die Begründung dieses Prinzips ergibt sich dann aus einer Verallgemeinerung der Argumentation dafür, dass M keine Ursache für W relativ zu {M, BM, W} sein kann, wenn BM eine Ursache für W relativ zu {M, BM, W} ist: Wenn BM eine Ursache für W relativ zu {M, BM, W} ist, dann kann M nur dann ebenfalls eine Ursache für W relativ zu {M, BM, W} sein, wenn es möglich ist, auf M zu intervenieren, während BM festgehalten wird. Genau das ist aber nicht möglich, wenn M auf BM nicht-kausal superveniert. Wenn BM also eine Ursache für W relativ zu {M, BM, W} ist und M auf BM superveniert, dann ist M keine Ursache für W relativ zu {M, BM, W}. Kontrapositorisch formuliert: Wenn M eine Ursache für W relativ zu {M, BM, W} ist und M auf BM superveniert, dann ist BM keine Ursache für W relativ zu {M, BM, W}.

Man beachte, dass dieses Exklusionsprinzip – wie auch alle anderen schon in Abschnitt 3.3.9. diskutierten Exklusionsprinzipien – neutral hinsichtlich der Frage ist, *welche* Ursache in einem gegebenen Fall ausgeschlossen wird. Eine Bevorzugung der subvenienten Variable bedarf also ebenso wie eine Bevorzugung der supervenierenden Variable weiterer Argumentation. Es ist keinesfalls offensichtlich, wie die Entscheidung hier aussehen sollte. Aus interventionistischer Sicht haben sowohl M als auch BM zunächst einen berechtigten Anspruch darauf, Ursache für W zu sein. Das ergibt sich aus einer Betrachtung der Variablenmengen {M, W} und {BM, W}. Jedoch können nicht beide Ursache für W relativ zu {M, BM, W} sein. Und es gibt innerhalb des Interventionismus, soweit ich ihn bis hier geschildert habe, kein Kriterium dafür, M oder BM als Ursache für W relativ zu {M, BM, W} zu bevorzugen.[194] Der Umgang mit der Variablenmenge {M, BM, W} bleibt also unklar, wenn statt der ursprünglichen Definitionen (DU) und (BU) die qualifizierten Definitionen (DU-Q) und (BU-Q) vorausgesetzt werden.

Siebtens: Wie schon mehrfach erwähnt sind alle in diesem Abschnitt entwickelten Ergebnisse auf Variablenmengen relativiert. Zudem wurden sie in Bezug

---

[194] Auf die Idee, M aufgrund von Überlegungen zu Proportionalität zu bevorzugen, gehe ich in Abschnitt 4.4.11. ein. Auf die Idee, BM aufgrund von Überlegungen zu kausaler Geschlossenheit zu bevorzugen, gehe ich in Abschnitt 4.4.9. ein.

auf Variablenmengen erreicht, die Variablen enthalten, die in Supervenienzbeziehungen stehen. Es ist daher recht naheliegend, das unerwünschte Ergebnis, dass weder M noch BM eine Ursache für W relativ zu {M, BM, W} ist, einfach dadurch zu vermeiden, dass man die Variablenmenge {M, BM, W} als unzulässige Variablenmenge von der Betrachtung ausschließt. Tatsächlich sind – neben Woodward selbst[195] – auch eine Reihe von anderen Autor*innen der Meinung, dass es sich bei {M, BM, W} einfach nicht um eine Variablenmenge handelt, auf die die interventionistischen Definitionen (DU) und (BU) angewendet werden sollten.[196] Was aber ist verkehrt an der Variablenmenge {M, BM, W}?

Vergleichen wir {M, BM, W} einmal mit einer anderen Variablenmenge {T, TS, N}, deren Variablen in Tabelle 4.9 bestimmt sind.[197]

Tabelle 4.9 Variablenmenge {T, TS, N}

Variable	Werte	Interpretation der Werte
T	t0 – t200	t0 = die Temperatur der Wasserprobe hat eine Temperatur von 0° C. t1 = die Temperatur der Wasserprobe hat eine Temperatur von 1° C. … t200 = die Temperatur der Wasserprobe hat eine Temperatur von 200° C.
TS	ts0, ts1	ts0 = die Temperatur der Wasserprobe liegt unter 100° C. ts1 = die Temperatur der Wasserprobe liegt bei oder über 100° C.
N	n0, n1	k0 = Die Nudeln werden nicht gar. k1 = Die Nudeln werden gar.

Auch in {T, TS, N} liegt eine Supervenienzbeziehung zwischen Variablen vor: TS superveniert auf T: Jede Änderung von TS geht notwendigerweise mit einer Änderung von T einher: Man kann beispielsweise nicht dafür sorgen, dass die Wasserprobe eine Temperatur von 100° C oder mehr annimmt (ts1), während man die Temperatur der Wasserprobe auf 90° C festhält (t90). Wie auch die Beziehung zwischen M und BM ist die Beziehung zwischen TS und T dabei nicht-kausal: Eine Änderung der Temperatur auf 101° C (t101) *verursacht* nicht, dass die Wasserprobe eine Temperatur von 100° C oder mehr annimmt (ts1), sondern kommt diesem Umstand in einem gewissen Sinne gleich. Die Beziehung zwischen T und TS ist nicht kausal, sondern logisch. Entsprechend ließe sich

[195] Woodward (2015) zieht jedoch Modifikationen der Definitionen einer direkten und beitragenden Ursache in Betracht.

[196] Vgl. Woodward (2017), Eronen (2012), Yang (2013).

[197] Vgl. für ein entsprechendes Beispiel Hoffmann-Kolss (2014: 59).

nun – parallel zu der obigen Überlegung mit Bezug auf M und BM – zeigen, dass weder T noch TS Ursachen für N relativ zu {T, TS, N} sein können.

Die Probleme im Zusammenhang mit Variablen, die in einer Supervenienzbeziehung stehen, sind also nicht spezifisch für den nicht-reduktiven Physikalismus und die ontologische Abhängigkeit mentaler Eigenschaften von ihren physischen Basen. Sie treten stattdessen zum Beispiel auch dann auf, wenn ein Modell zwei Variablen enthält, die wie T und TS in einem Schwellenwert-Verhältnis stehen (so dass eine Variable genau dann einen bestimmten Wert annimmt, wenn die Werte der anderen Variable eine bestimmte Schwelle überschreiten). Ähnliche Probleme ergeben sich, wenn die Werte verschiedener Variablen Sachverhalte repräsentieren, die durch arithmetische oder Boole'sche Operationen auseinander konstruiert werden können: Ein bekanntes Beispiel ist hier der Zusammenhang zwischen einer ersten Variable LC, die den LDL-Cholesterinspiegel (‚low density lipoprotein') einer Person anzeigt, einer zweiten Variable HC, die den HDL-Cholesterinspiegel (‚high density lipoprotein') einer Person anzeigt, und einer dritten Variable GC, die den Gesamt-Cholesterinspiegel einer Person anzeigt, wobei der Gesamt-Cholesterinspiegel als die Summe aus dem HDL-Cholesterinspiegel und dem LDL-Cholesterinspiegel bestimmt ist. Auch hier ist es offenbar nicht möglich, den LDL-Cholesterinspiegel oder den HDL-Cholesterinspiegel zu verändern, während der Gesamt-Cholesterinspiegel festgehalten wird. Entsprechend könnte man (irreführenderweise) folgern, dass der LDL-Cholesterinspiegel keine Ursache für Herzkrankheiten ist relativ zu einer Variablenmenge, die auch eine Variable enthält, die den Gesamt-Cholesterinspiegel repräsentiert.[198]

Die Betrachtung von Variablenmengen, die Variablen enthalten, die Sachverhalte, Eigenschaften oder Ereignisse repräsentieren, die in Beziehungen der ontologischen Abhängigkeit oder in logischen Beziehungen stehen, führt also in vielen Fällen zu falschen oder irreführenden (relativierten) Kausalaussagen. Vor diesem Hintergrund erscheint das Problem mit der Variablenmenge {M, BM, W} als ein Spezialfall eines allgemeineren Problems, das sich aus der Betrachtung ungeeigneter Variablenmengen ergibt. Dies legt nahe, dass Interventionist*innen ganz unabhängig von einer Festlegung auf eine bestimmte metaphysische Hintergrundtheorie wie den nicht-reduktiven Physikalismus eine Einschränkung für die Zulässigkeit von Variablenmengen einführen sollten: Die Variablen in einer

[198] Vgl. Woodward (2015: 327–332). Auch aus begrifflichen Zusammenhängen zwischen repräsentierten Sachverhalten ergeben sich entsprechende Probleme: Hier wäre ein Beispiel eine Variablenmenge, in der eine Variable anzeigt, ob eine Person ein Junggeselle ist und eine andere Variable anzeigt, ob die Person unverheiratet ist.

Variablenmenge, auf die die Definitionen der direkten und beitragenden Ursache angewendet werden, sollten auf eine noch näher zu spezifizierende Weise unabhängig voneinander sein.

Woodward führt in diesem Kontext ein Kriterium der unabhängigen Fixierbarkeit ein:

> (IF): a set of variables V satisfies independent fixability of values if and only if for each value it is possible for a variable to take individually, it is possible (that is, "possible" in terms of their assumed definitional, logical, mathematical, mereological or supervenience relations) to set the variable to that value via an intervention, concurrently with each of the other variables in V also being set to any of its individually possible values by independent interventions.[199]

Das Kriterium der unabhängigen Fixierbarkeit verlangt auf den Kern heruntergebrochen, dass alle Kombinationen von Werten von Variablen in einer Variablenmenge möglich sein sollten. Sobald eine Kombination nicht möglich ist, kann die entsprechende Kombination auch nicht mehr durch Interventionen herbeigeführt werden.[200] Offenbar erfüllen die Variablenmengen {M, BM, W}, {T, TS, N} und {LC, HC, TC} das Kriterium der unabhängigen Fixierbarkeit nicht. In allen Fällen gibt es Kombinationen von Variablenwerten, die unmöglich sind und daher nicht durch Interventionen herbeigeführt werden können. Es handelt sich also nicht um *unabhängig fixierbare Variablenmengen.*

Alle bis hierher diskutierten unerwünschten Ergebnisse lassen sich nun dadurch vermeiden, dass man die Anwendungsreichweite des Interventionismus auf unabhängig fixierbare Variablenmengen beschränkt. Erstens stehen M und BM laut dem Interventionismus dann nicht in einem kausalen Verhältnis zueinander: Es gibt keine (zulässige) Variablenmenge, relativ zu der M eine direkte oder beitragende Ursache für BM ist, weil die Variablenmenge {M, BM} von der Betrachtung ausgeschlossen ist. Zweitens ist die Annahme der Monotonie nicht gefährdet: Wir können davon ausgehen, dass M eine Ursache für W relativ zu {M, W} ist, ohne zu akzeptieren, dass es eine (zulässige) Erweiterung von {M, W} gibt, relativ zu der M keine beitragende Ursache für W ist. Denn die Variablenmenge {M, BM, W} ist als unzulässig von der Betrachtung ausgeschlossen. Drittens können wir die Feststellung, dass weder M noch BM relativ zu {M, BM,

---

[199] Woodward (2015: 316).

[200] Ob Woodwards Formulierung der unabhängigen Fixierbarkeit noch mehr verlangt als die Möglichkeit aller Kombinationen von Variablenwerten, hängt davon ab, ob es Kombinationen von Variablenwerten gibt, die zwar eigentlich im relevanten Sinne möglich sind, die aber nicht *durch Interventionen herbeigeführt* werden können. Da mir hiervon aber im vorliegenden Kontext nichts abzuhängen scheint, übergehe ich diese Frage.

W} Ursache für W sind, nun einordnen: Diese Feststellung hat keinerlei Konsequenzen für die tatsächlichen Kausalverhältnisse zwischen M, BM und W, weil sie eine unzulässige Variablenmenge betrifft.[201]

Wie schon angedeutet ist das Kriterium der unabhängigen Fixierbarkeit dabei unabhängig vom nicht-reduktiven Physikalismus motiviert. Es handelt sich daher nicht um eine ad hoc-Modifikation des Interventionismus in Anbetracht seiner Probleme, sich in die Metaphysik des nicht-reduktiven Physikalismus einzupassen. Vielmehr ist eine solche Modifikation oder Ergänzung des Interventionismus in jedem Fall und unabhängig von vorgängigen metaphysischen Annahmen nötig, um zum Beispiel die Probleme mit den Variablenmengen {TS, T, N} und {HC, LC, GC} in den Griff zu bekommen.

Die Beschränkung auf unabhängig fixierbare Variablenmengen wird also im Folgenden vorausgesetzt. Dies löst jedoch nicht alle Probleme, die sich aus einer Verbindung des nicht-reduktiven Physikalismus mit dem Interventionismus ergeben. Ein anderes, eng verwandtes Problem ergibt sich aus Woodwards Definition einer Intervention. Im folgenden Abschnitt schildere ich zunächst, wie sich aus dieser Definition ein interventionistisches Exklusionsprinzip ergibt. Daran anschließend diskutiere ich, ob sich von diesem Exklusionsprinzip ausgehend ein starkes interventionistisches Exklusionsargument entwickeln lässt.

4.4.7 Interventionen und interventionistische Exklusion

Die im vorigen Abschnitt behandelten Probleme ergeben sich aus der in den Definitionen der direkten und beitragenden Verursachung enthaltenen Forderung des Festhaltens anderer Variablen. Ich habe in meiner Diskussion dabei vorausgesetzt, dass Interventionen auf die in Frage stehenden Variablen möglich sind. In diesem Abschnitt wird nun ein Problem diskutiert, das sich für die Möglichkeit von Interventionen ergibt. Ausgangspunkt ist dabei wieder die Beobachtung, dass der nicht-reduktive Physikalismus bestimmte Kombinationen von Variablenwerten unmöglich macht und daher zu Supervenienz-Beziehungen zwischen Variablen führt.[202]

Wie oben geschildert versteht Woodward Interventionen auf X als Werte von Interventionsvariablen, die Werte von X aktual verursachen. Hieraus folgt, dass

[201] Hoffmann-Kolss (2021) zeigt, dass es Fälle gibt, die durch die Forderung der unabhängigen Fixierbarkeit nicht in den Griff zu bekommen sind.

[202] Vgl. für die folgende Argumentation Baumgartner (2009) und Baumgartner (2010).

es eine Intervention auf X hinsichtlich Y natürlich nur dann geben kann, wenn es überhaupt eine Interventionsvariable für X hinsichtlich Y gibt.

Zur Erinnerung hier noch einmal Woodwards Definition einer Interventionsvariablen:

> **(IV)** Eine Variable *I* ist eine Interventionsvariable auf eine Variable *X* hinsichtlich einer Variable *Y* genau dann, wenn gilt:
>
> (I1) *I* verursacht *X*.
>
> (I2) *I* ist ein Umschalter für *X*. Das heißt, bestimmte Werte von *I* sind so beschaffen, dass gilt: Wenn *I* diese Werte annimmt, hängt *X* nicht von den Werten anderer Ursachen von *X* ab, sondern hängt stattdessen nur vom Wert von *I* ab.
>
> (I3) Jeder kausale Pfad von *I* zu *Y* geht durch *X*. Das heißt, *I* ist keine direkte Ursache von *Y* und ist keine Ursache von Ursachen von *Y*, die distinkt von *X* sind, außer jenen Ursachen von *Y*, die in den gerichteten Pfad von *I* über *X* zu *Y* eingebaut sind; also (a) jenen Ursachen von *Y*, die Wirkungen von *X* sind und (b) jenen Ursachen von *Y*, die auf einem kausalen Pfad zwischen *I* und *X* liegen und die keine von *X* unabhängigen Auswirkungen auf *Y* haben.
>
> (I4) *I* ist statistisch unabhängig von jeder Variablen *Z*, die *Y* verursacht und die auf einem kausalen Pfad liegt, der nicht durch *X* geht.[203]

In Woodwards Definition einer Interventionsvariablen (IV) wird einerseits verlangt, dass Interventionsvariablen für eine Variable X hinsichtlich einer Variablen Y Ursachen für X sind (I1 und I2). Andererseits wird in (IV) auch verlangt, dass Interventionsvariablen für eine Variable X hinsichtlich einer Variable Y auf bestimmte Weise unabhängig von anderen Ursachen für Y sind: Interventionsvariablen dürfen erstens, wie in Bedingung (I3) spezifiziert, Y nur über Pfade verursachen, die X einschließen und sie müssen zweitens, wie in (I4) spezifiziert, auch statistisch unabhängig von allen Ursachen von Y sein, die nicht von X verursacht werden und die X nicht verursachen.

Wenden wir diese Forderungen auf die Variable M an und setzen dabei voraus, dass M auf BM nicht-kausal superveniert. Es kann nun gezeigt werden, dass es keine Interventionsvariablen für M hinsichtlich Wirkungen von BM geben kann:

Gehen wir davon aus, dass BM eine direkte Ursache für W relativ zu {BM, W} ist. BM ist also auch simpliciter eine Ursache für W. Gehen wir weiter davon aus, dass BM keine Ursache oder Wirkung von M ist. Diese Annahme ist durch die Forderung des nicht-reduktiven Physikalismus gedeckt, dass M *nicht-kausal*

[203] Vgl. Woodward (2003: 98).

auf BM superveniert.[204] Hieraus folgt nach (IV4), dass Interventionsvariablen auf M hinsichtlich W von BM statistisch unabhängig sein müssen: Denn BM ist eine Ursache von W, die nicht auf einem kausalen Pfad von M zu W liegt. Zudem folgt nach (I3), dass Interventionsvariablen auf M hinsichtlich W keine Ursachen für BM sein dürfen. Denn BM ist eine Ursache für W, die distinkt von X ist und nicht in den kausalen Pfad von I über X zu Y eingebaut ist.

(IV) verlangt jedoch auch, dass Interventionsvariablen auf M hinsichtlich W Ursachen für M sind. Nun gibt es jedoch keine Ursache für M, die statistisch unabhängig von BM ist.[205] Und es gibt keine Ursache für M, die nicht zugleich auch Ursache für BM ist.[206] Jede Ursache für M ist statistisch abhängig von BM und verursacht BM. Dies liegt einfach daran, dass M auf BM superveniert. Eine Änderung von M geht notwendig immer mit einer Änderung von BM einher. *Also gibt es keine Interventionsvariable auf M hinsichtlich W*.

Man beachte, dass dieses Ergebnis nur unter der Voraussetzung erreicht werden kann, dass BM eine Ursache für W (relativ zu {BM, W}) ist. Wenn BM keine Ursache für W ist, müssen Interventionsvariablen auf M hinsichtlich W auch nicht statistisch unabhängig von BM sein und sie dürfen BM verursachen. Wenn BM keine Ursache für W ist, steht der Existenz von Interventionsvariablen auf M hinsichtlich W also auch nichts im Wege. Unter Berücksichtigung hiervon kann das Ergebnis also wie folgt ausgedrückt werden: *Wenn BM eine Ursache für W ist, dann gibt es keine Interventionsvariable auf M hinsichtlich W*.

---

[204] Man beachte auch, dass die Variablenmenge {M, BM} und alle ihre Erweiterungen aufgrund des Kriteriums der unabhängigen Fixierbarkeit ausgeschlossen sind. Schon deshalb kann es keine zulässige Variablenmenge geben, relativ zu der M eine direkte oder beitragende Ursache für BM ist oder BM eine direkte oder beitragende Ursache für M ist. Tatsächlich scheint mir die vorliegende Rekonstruktion von Baumgartners Argument also sogar vorauszusetzen, dass ein Kriterium der unabhängigen Fixierbarkeit (oder etwas ähnliches) angewendet wird. Baumgartners eigene Rekonstruktionen machen dennoch Gebrauch von Variablenmengen, die das Kriterium der unabhängigen Fixierbarkeit verletzen.

[205] Man muss hier eigentlich qualifizieren: Es gibt keine totalen Ursachen für M, die statistisch unabhängig von BM sind. Denn Ursachen für M, die keine totalen Ursachen für M sind, müssen von M nicht statistisch abhängig sein. Sie müssen daher auch von BM nicht statistisch abhängig sein. Da Interventionsvariablen aber Umschalter für M sein müssen (I2), ist es ohnehin ausgeschlossen, dass sie keine totalen Ursachen für M sind.

[206] Auch hier müsste man glaube ich qualifizieren bzw. genauer argumentieren: Zwar ist jede Intervention auf I hinsichtlich M, die M verändert, auch eine Intervention auf I hinsichtlich M, die BM verändert. Folgt aber daraus, dass es eine Intervention auf I hinsichtlich M gibt, die BM verändert, zugleich, dass es eine Intervention auf I *hinsichtlich BM* gibt, die BM verändert? Es könnte prinzipiell schließlich sein, dass BM Ursachen hat, die keine Ursachen für M sind und dass jede Intervention auf I von diesen zusätzlichen Ursachen von BM abhängt.

Wenn es keine Interventionsvariable auf M hinsichtlich W gibt, dann kann M auch keine Ursache für W sein. Für diesen Schritt kann man sich wieder auf Woodwards Definitionen der totalen, direkten und beitragenden Verursachung berufen. In jeder dieser Definitionen wird verlangt, dass eine Variable X nur dann eine Ursache für eine Variable Y sein kann, wenn es eine mögliche Intervention auf X hinsichtlich Y gibt. Wenn es aber schon keine Interventions*variable* auf M hinsichtlich W gibt, dann gibt es natürlich auch keine mögliche Intervention auf M hinsichtlich W. Also folgt aus dem Ergebnis, *dass M keine Ursache für W ist, wenn BM eine Ursache für W ist*.

Das Ergebnis lässt sich wieder verallgemeinern: Wenn eine Variable X auf einer Variable Y nicht-kausal superveniert und Y eine Variable Z verursacht, dann kann X keine Ursache für Z sein. Denn es gibt dann keine Interventionsvariable für X hinsichtlich Z.[207] In dieser allgemeinen Formulierung kommt das Ergebnis einem interventionistischen Exklusionsprinzip gleich:

> **Das interventionistische Exklusionsprinzip (IE)**: Für alle Variablen X, Y, Z: Wenn X nicht-kausal auf Y superveniert und Y eine Ursache für Z ist, dann ist X keine Ursache für Z.

Ein allgemeines Argument für das interventionistische Exklusionsprinzip kann wie folgt konstruiert werden:

Argument für das interventionistische Exklusionsprinzip:

> **(IE1)** Für alle Variablen X, Y, Z: Wenn es eine Interventionsvariable I auf X hinsichtlich Z gibt und Y eine Ursache für Z ist und X keine Ursache für Y ist und Y keine Ursache für X ist, dann gibt es eine Variable I, die X verursacht und die statistisch unabhängig von Y ist. (Prämisse)

> **(IE2)** Für alle Variablen X, Y: Wenn X nicht-kausal auf Y superveniert, dann ist X keine Ursache für Y und Y ist keine Ursache für X. (Prämisse)

> **(IE3)** Für alle Variablen X, Y: Wenn X nicht-kausal auf Y superveniert, dann gibt es keine Variable I, die X verursacht und die statistisch unabhängig von Y ist. (Prämisse)

[207] Diese Überlegung setzt natürlich voraus, dass der Begriff einer Interventionsvariablen nicht relativ zu seiner Variablenmenge definiert wird. Diese Voraussetzung diskutiere ich weiter unten noch.

(IE4) Also: Für alle X, Y, Z: Wenn X auf Y nicht-kausal superveniert und Y eine Ursache für Z ist, dann gibt es keine Interventionsvariable auf X hinsichtlich Z. (aus IE1, IE2 und IE3)

(IE5) Für alle Variablen X, Z: Wenn es keine Interventionsvariable auf X hinsichtlich Z gibt, dann ist X keine Ursache für Z. (Prämisse)

(IEK) Also: Für alle X,Y,Z: Wenn X nicht-kausal auf Y superveniert und Y eine Ursache für Z ist, dann ist X keine Ursache für Z (aus IE4 und IE5).

(IE1) ergibt sich dabei aus der Definition einer Interventionsvariablen. (IE2) legt fest, dass Variablen, die in einer nicht-kausalen Supervenienzbeziehung stehen, einander nicht verursachen. (IE3) legt fest, dass jede Ursache einer supervenienten Variable statistisch abhängig von der zugehörigen subvenienten Variable ist. Dies ergibt sich daraus, dass es im Allgemeinen nicht möglich ist, eine supervenierende Variable zu ändern, ohne auch die zugehörige subvenierende Variable zu ändern. (IE4) folgt logisch aus (IE1), (IE2) und (IE3). (IE5) ergibt sich aus Woodwards Definitionen kausaler Begriffe. (IEK) folgt aus (IE4) und (IE5).

Wieder sind eine Reihe von Kommentaren zu dem interventionistischen Exklusionsprinzip und der Argumentation hierfür angebracht:

Erstens: Das interventionistische Exklusionsprinzip enthält keine Relativierungen auf Variablenmengen. Es betrifft stattdessen unmittelbar simpliciter Verursachung. Dies unterscheidet die vorliegende Argumentation in entscheidender Hinsicht von der Argumentation aus Abschnitt 4.4.6.: Während es in 4.4.6. immer um relativierte Kausalaussagen ging, geht es hier von vornherein um nicht-relativierte Kausalaussagen.

Das interventionistische Exklusionsprinzip kommt ohne eine Relativierung auf Variablenmengen aus, weil es von der Definition einer Interventionsvariablen ausgeht. Die Definition einer Interventionsvariablen aber ist nicht auf Variablenmengen relativiert und macht Gebrauch von de-relativierten Kausalaussagen. Interventionsvariablen auf M hinsichtlich W müssen statistisch unabhängig von BM sein, insofern BM eine *de-relativierte* Ursache für W ist, die im *de-relativierten Sinn von Verursachung* nicht kausal mit M verknüpft ist.

Dass ein solches de-relativiertes Verständnis von Interventionsvariablen (auf den ersten Blick) auch nötig ist, verdeutlicht das folgende Beispiel von Michael Strevens: Nehmen wir an, dass das Essen salziger Nahrung (S) eine gemeinsame Ursache für die Menge von konsumiertem Wasser (W) und das Entwickeln eines Herzleidens (H) ist, so dass es zwar eine Korrelation, aber keine kausale Beziehung zwischen der

Menge von konsumierten Wasser und dem Entwickeln eines Herzleidens gibt. Relativ zur Variablenmenge {S,W,H} würde sich dann folgende Struktur gemeinsamer Verursachung ergeben (Abbildung 4.7):

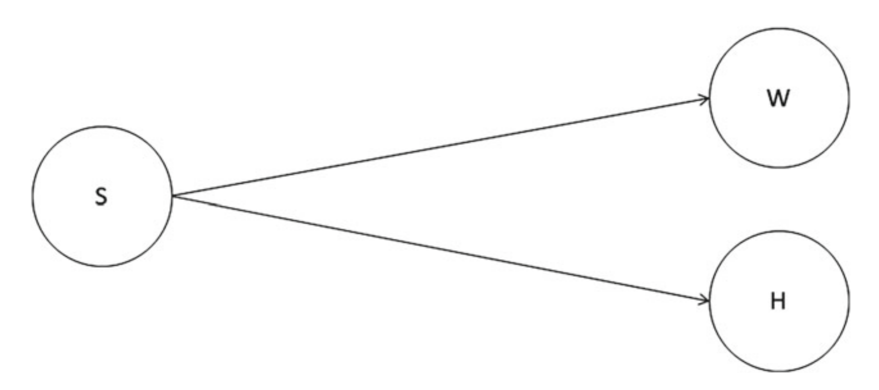


Abbildung 4.7 kausaler Graph für die Variablenmenge {S, W, H}

Strevens meint nun, dass die Entfernung der gemeinsamen Ursache S aus der Variablenmenge V nach Woodwards Definitionen zur Konsequenz hat, dass W als Ursache von H klassifiziert wird. W ist also zwar keine beitragende Ursache von H relativ zu {S,W,H}, jedoch ist W eine beitragende Ursache von H relativ zu {W,H}.

Wenn Strevens Argumentation funktioniert, hat dies für den Interventionismus katastrophale Konsequenzen: Da es eine Variablenmenge gibt, relativ zu der W Ursache von H ist, müssten wir schließen dass W simpliciter Ursache von H ist. Jedoch ist W offensichtlich nicht Ursache von H – die tatsächliche kausale Struktur wird stattdessen in der Variablenmenge {S, W, H} deutlich: W und H haben eine gemeinsame Ursache. Nach Strevens kann der Interventionismus also nicht einmal mit einfachen Fällen gemeinsamer Verursachung umgehen. Zudem würde Strevens Argumentation die Annahme der Monotonie relativierter Verursachung widerlegen: W ist eine Ursache für H relativ zu {W,H}, aber nicht relativ zur Erweiterung {S,W,H}.

Strevens argumentiert wie folgt dafür, dass W eine Ursache für H relativ zu {W,H} ist:

> Because Woodward's definition of an intervention is implicitly relativized to the variables in a network, increasing the amount of bottled water you drink by increasing your consumption of salty foods will count as an intervention relative to the salt-free

> network (due to the invisibility, within the network, of the “side effects” of the salty strategy for drinking more).[208]

Strevens geht also davon aus, dass die Definitionen einer Intervention und einer Interventionsvariablen implizit ebenfalls auf eine Variablenmenge relativiert sind. Eine Intervention auf W hinsichtlich H *relativ zu {W,H}* muss daher nicht von S statistisch unabhängig sein, weil S nicht in {W,H} enthalten ist. Daher gibt es Interventionen auf W hinsichtlich H *relativ zu {W,H}*, die H verändern – nämlich solche, die auch S verändern.

Woodward reagiert auf diesen Einwand, indem er betont, dass seine Definitionen einer Intervention und einer Interventionsvariablen *nicht* auf Variablenmengen relativiert sind:

> [A] look at IN makes it clear that there is no explicit or obvious relativization to a variable set of the sort that Strevens has in mind. In particular, IN is formulated in terms of requirements that concern the relationship between the intervention variable I and “other (contributing) causes” of Y, the putative effect variable, and not in terms of the relationship between I and the other causes of Y that are in some particular variable set V or that are known to the experimenter. In other words, the intervention must be uncorrelated with all potential confounders, not just with all confounders that happen to be in some variable set such as the one we use to describe the system in which the intervention occurs. Thus, contrary to what Strevens claims, to count as an intervention on W with respect to H, the manipulation of water consumption must not be correlated with S. There is no such thing, from my point of view, as the manipulation counting as an intervention with respect to W and H but not with respect to W, H and S. Indeed, it was precisely to avoid consequences like those described by Strevens […] that I very deliberately elected not to relativize the notion of an intervention to a variable set.[209]

Eine Intervention auf W muss also von jeder Ursache von H statistisch unabhängig sein – ganz gleich, ob diese Ursache in der betrachteten Variablenmenge {W,H} repräsentiert ist oder nicht. Da auch S eine Ursache für H (z. B. relativ zu {S, H} oder {S, W, H} und daher simpliciter) ist, muss die Intervention auf W also auch statistisch unabhängig von S sein. Die von Strevens in den Blick genommene Veränderung von W, die auch eine Änderung von S zur Folge hat, zählt also nicht als Intervention auf W hinsichtlich H. Daher müssen wir nicht schließen, dass W eine Ursache von H relativ zu {W, H} ist. Die unerwünschten Konsequenzen sind vermieden.

---

[208] Strevens (2007: 243).

[209] Woodward (2008b: 21).

Der Austausch zwischen Strevens und Woodward verdeutlicht, dass die Definitionen einer Intervention und einer Interventionsvariable nicht auf eine Variablenmenge relativiert werden sollten.[210] Stattdessen beziehen sich die Verwendungen kausaler Begriffe in diesen Definitionen immer auf de-relativierte Kausalaussagen. Entsprechend muss eine Intervention auf eine Variable X hinsichtlich einer Variable Y schon dann statistisch unabhängig von einer pfadunabhängigen Variable Z sein, wenn es überhaupt eine Variablenmenge gibt, relativ zu der Z eine Ursache für Y ist.

Um im Argument für das interventionistische Exklusionsprinzip daher zu etablieren, dass Interventionsvariablen auf M hinsichtlich W statistisch unabhängig von BM sein müssen, ist es nur entscheidend, dass es *eine Variablenmenge gibt, relativ zu der BM eine Ursache für W ist*. Diese Variablenmenge muss M nicht enthalten und muss also auch nicht das Kriterium der unabhängigen Fixierbarkeit verletzen. Es bietet sich hier die Variablenmenge {BM, W} an: Solange BM eine Ursache für W relativ zu {BM, W} ist, müssen Interventionsvariablen auf M hinsichtlich W unabhängig von BM sein.

Zweitens: Das Kriterium der unabhängigen Fixierbarkeit ist deshalb keine Hilfe dabei, das interventionistische Exklusionsprinzip zu vermeiden. Denn es wird nirgendwo vorausgesetzt, dass man Variablenmengen betrachten sollte, die Variablen enthalten, die in Supervenienzbeziehungen stehen. Es reicht stattdessen, Variablenmengen wie {BM, W} zu betrachten, die BM und die vermeintlichen Wirkungen von M enthalten. Wenn BM relativ zu {BM, W} eine direkte Ursache für M ist, dann kann M relativ zu *keiner* Variablenmenge eine direkte oder beitragende Ursache für W sein.[211]

Ganz im Gegenteil ist das Kriterium der unabhängigen Fixierbarkeit der vorliegenden Argumentation für das interventionistische Exklusionsprinzip sogar zuträglich: Ohne ein solches Kriterium müsste man (prima facie – ohne eine

[210] Dies setzt allerdings voraus, dass das von Strevens präsentierte Problem nur durch Woodwards Erwiderung gelöst werden kann. Sowohl McCain (2015) als auch Statham (2017) weisen jedoch darauf hin, dass Strevens Beispiel schon aus einem anderen Grund nicht funktioniert: Die von Strevens in den Blick genommene Intervention durch Veränderung von S wäre selbst eine Ursache für Y und würde daher (I3) verletzen. Dies ist schon dann der Fall, wenn man zunächst von der Variablenmenge {W, H} ausgeht.

Zudem zeigt Strevens Argumentation im Erfolgsfall nur, dass die Definition einer Interventionsvariablen I auf X hinsichtlich Y nicht in dem naiven Sinn auf eine Variablenmenge V relativiert sein sollte, dass nur die Unabhängigkeit von I von den Ursachen von Y in V verlangt wird. Sie zeigt jedoch nicht, dass komplexere Formen der Relativierung, die auch Erweiterungen von V in Betracht ziehen, nicht funktionieren. Vgl. für eine solche komplexere relativierte Definition einer Interventionsvariablen Weslake (2011).

[211] Vgl. Baumgartner (2013b: 10–11).

andere Lösung für das Problem) Variablen, die in Supervenienzbeziehungen stehen, nämlich als kausal miteinander verknüpft betrachten. Man müsste insbesondere schließen, dass M eine Ursache für BM relativ zu {M, BM} ist. Dann aber würde (IV) nicht mehr verlangen, dass eine Interventionsvariable für M hinsichtlich W statistisch unabhängig von BM sein muss. BM läge dann schließlich auf einem kausalen Pfad von M zu W.[212] Das Argument setzt also sogar voraus, dass Variablenmengen, die das Kriterium der unabhängigen Fixierbarkeit nicht erfüllen, unzulässig sind.

Drittens: Das interventionistische Exklusionsprinzip ist – wie auch alle anderen, schon in Abschnitt 3.3.9. behandelten Exklusionsprinzipien – zunächst neutral gegenüber der Frage, welche der beiden in Frage kommenden Ursachen ausgeschlossen wird. Es legt nicht fest, dass M nicht als Ursache für W zählt, sondern nur, dass M nicht als Ursache für W zählt, wenn BM als Ursache für W zählt. Wenn BM hingegen nicht als Ursache für W zählt, kann M problemlos als Ursache für W zählen. Das Exklusionsprinzip stellt uns also vor eine Wahl: Entweder, M verursacht W. Dann kann BM keine Ursache für W sein. Oder BM verursacht W. Dann kann M keine Ursache für W sein. Aus dem Interventionismus selbst lässt sich dabei zunächst keine Entscheidung herbeiführen. Betrachtet man die Variablenmenge {M, W}, sieht M wie eine Ursache für W aus. Und betrachtet man die Variablenmenge {BM, W}, sieht BM wie eine Ursache für W aus. Gegeben das interventionistische Exklusionsprinzip ist dieser erste Anschein jedoch trügerisch. Eine der beiden Variablenmengen führt zu einem Fehlurteil. Welche der beiden Variablenmengen allerdings zu einem Fehlurteil verleitet, kann nicht allein auf der Grundlage von Interventionen entschieden werden. Was vielmehr in Frage steht, ist, welche Manipulationen überhaupt als Interventionen zählen können. Und um dies zu entscheiden, müssen bereits bestimmte Annahmen über die kausalen Beziehungen zwischen M, BM und W vorausgesetzt werden.

Viertens: Das interventionistische Exklusionsprinzip kann durch bestimmte Modifikationen der Definition einer Interventionsvariable vermieden werden.[213] Auf diese Strategie gehe ich in Abschnitt 4.4.10 noch einmal ein. Zunächst möchte ich im folgenden Abschnitt jedoch die Frage behandeln, ob sich aus dem interventionistischen Exklusions*prinzip* ein plausibles interventionistisches Exklusions*argument* ergibt.

[212] Vgl. hierzu auch Kroedel (2019: 149).

[213] Vgl. Woodward (2017: 333–334) und Weslake (2011).

4.4.8 Das interventionistische Exklusionsargument

Klarerweise lässt sich auf der Grundlage des interventionistischen Exklusionsprinzips das folgende Exklusionsargument konstruieren:

Das interventionistische Exklusionsargument

> **(IEA1)** Wenn BM eine Ursache für W ist und M auf BM nicht-kausal superveniert, dann ist M keine Ursache für W.
>
> **(IEA2)** M superveniert nicht-kausal auf BM.
>
> **(IEA3)** BM ist eine Ursache für W.
>
> **(IEAK)** Also: M ist keine Ursache für W.

Dieses Argument ist gültig. Es bleibt jedoch die Frage, ob seine Prämissen wahr sind, ob sie nicht-zirkulär begründet werden können und ob sie von Vertreter*innn des nicht-reduktiven Physikalismus akzeptiert werden müssen.

(IEA1) ist einfach eine Instanz des interventionistischen Exklusionsprinzips, dessen Begründung ich in Abschnitt 4.4.8. besprochen habe: Wenn M auf BM superveniert und BM W verursacht, dann sind Interventionen auf M hinsichtlich W unmöglich. M kann daher dann keine Ursache für W sein. Die Ablehnung dieser Prämisse durch Modifikationen der Definition einer Interventionsvariable bespreche ich in Abschnitt 4.4.10.

(IEA2) scheint aus Sicht des nicht-reduktiven Physikalismus zunächst zwingend zu sein: Wie in Abschnitt 4.4.5. geschildert, ergibt sich aus dem nicht-reduktiven Physikalismus, dass bestimme Kombinationen von Werten von M und BM unmöglich sind und dass jeder Wert von M von einem Wert von BM necessitiert wird. Zudem ist mit dem nicht-reduktiven Physikalismus verbunden, dass Instantiierungen von **M** nicht von den Instantiierungen der Basen von **M** (**BM1** und **BM2**) verursacht werden.

Zugleich ist jedoch zu beachten, dass (IEA2) im Rahmen des Interventionismus nur dann plausibel wird, wenn der Interventionismus um ein Mittel erweitert wird, um Beziehungen der nicht-kausalen ontologischen Abhängigkeit von kausalen Beziehungen zu unterscheiden. Wie in Abschnitt 4.4.7. erläutert, kann dies zum Beispiel dadurch erreicht werden, dass man eine Einschränkung zulässiger Variablenmengen einführt: Zulässig sind dann nur solche Variablenmengen, die das

Kriterium der unabhängigen Fixierbarkeit erfüllen. Gegeben eine solche Einschränkung sind die Variablenmenge {M, BM} und alle ihre Erweiterungen unzulässig und die Beobachtung, dass Interventionen auf M zu Änderungen in BM führen, hat keine Konsequenzen für die Frage, ob M eine Ursache für BM ist. In Abwesenheit einer solchen Modifikation ist (IEA2) jedoch falsch. Das Problem der Kombination aus nicht-reduktivem Physikalismus und Interventionismus besteht dann nicht darin, dass **M** keine physischen Wirkungen haben kann, sondern darin, dass **M** seine Basen verursacht – was laut dem nicht-reduktiven Physikalismus ausgeschlossen sein sollte.[214]

Wie steht es um (IEA3)? Ich argumentiere im Folgenden dafür, dass nicht-reduktive Physikalist*innen die Annahme, dass BM eine Ursache für W ist, ablehnen können. Dies eröffnet die prinzipielle Möglichkeit für nicht-reduktive Physikalist*innen, das interventionistische Exklusionsprinzip zu akzeptieren. Ob diese Option aber zu einer zufriedenstellenden interventionistischen Konzeption mentaler Verursachung im nicht-reduktiven Physikalismus führen kann, hängt auch davon ab, ob Interventionist*innen gute Gründe dafür anführen können, M als Ursache für W gegenüber BM zu bevorzugen. Dies wiederum ist davon abhängig, ob Interventionist*innen ein starkes Proportionalitätsprinzip in ihre Theorie einbauen können. Auf diese Frage gehe ich erst in Abschnitt 4.4.10 genauer ein.

Die naheliegendste Möglichkeit, (IEA3) im interventionistischen Exklusionsargument zu motivieren, besteht in einer Berufung auf die These der kausalen Geschlossenheit. Eine solche Begründung hat den Vorteil, dass sie sich auf eine These beruft, die üblicherweise als Kernthese des nicht-reduktiven Physikalismus verstanden wird. Da nicht-reduktive Physikalist*innen also auf die These der kausalen Geschlossenheit festgelegt sind, können sie die Annahme, dass BM eine Ursache für W ist, dann nicht zurückweisen.[215]

Meine Kommentare zur These der kausalen Geschlossenheit in Abschnitt 3.2. ziehen jedoch in Zweifel, ob Prämisse (IEA3) angemessen durch die These der kausalen Geschlossenheit begründet werden kann:

Erstens sind nicht-reduktive Physikalist*innen in erster Linie auf eine These der nomologischen Geschlossenheit festgelegt. Ob hieraus auch eine These der kausalen Geschlossenheit folgt, hängt von der vorausgesetzten Theorie der Kausalität ab.

[214] Vgl. Kroedel (2019).

[215] Baumgartner beruft sich auf eine solche Verbindung zwischen der Prämisse, dass BM eine Ursache für W ist, und der These der kausalen Geschlossenheit (vgl. Baumgartner (2013b: 3, 8). Er führt jedoch auch an, dass die Prämisse schwächer als die These der kausalen Geschlossenheit sei (vgl. Baumgartner (2009: 171)). Ich argumentiere dafür, dass die Prämisse nicht schwächer als die These der kausalen Geschlossenheit ist, sondern logisch unabhängig von dieser These.

Zudem bezieht sich die These der kausalen Geschlossenheit auf den Begriff der hinreichenden Verursachung. Hinreichende Verursachung impliziert jedoch nicht kausale Abhängigkeit.[216] Prämisse (IEA3) betrifft aber kausale Abhängigkeit. Dies macht es fraglich, ob Prämisse (IEA3) unter Berufung auf die These der kausalen Geschlossenheit begründet werden kann.

Woodward schreibt zur Verbindung zwischen hinreichenden Bedingungen und interventionistischer Kausalität das Folgende:

> [Interventionism] yields a conception of causation that is importantly distinct from conceptions based on the idea that causes are (or are non-redundant "parts of" or "conjuncts in") sufficient conditions for their effects, which seems to be the conception often adopted in discussions of the exclusion argument. In particular, it is entirely possible for various individual token realizations (described in some "lower level" science) of some upper level variable X to be sufficient for individual token realizations of some upper level variable Y (also described in terms of some lower level science) and yet for it not to be the case that intervening on X is a means for changing Y.[217]

Selbst wenn jeder Wert einer Variable X nomologisch hinreichend dafür ist, dass Y einen bestimmten Wert annimmt, folgt hieraus noch nicht, dass eine Intervention auf X zu einer Änderung von Y führt. Insbesondere bricht die Folgerung dann zusammen, wenn die verschiedenen Werte von X für denselben Wert von Y nomologisch hinreichend sind. Sind zum Beispiel sowohl die Einnahme des Placebos als auch die Einnahme des Medikaments hinreichend dafür, dass die Krankheit geheilt wird, kann eine Intervention auf die Variable MED, die repräsentiert, ob ein Patient das Placebo oder das Medikament einnimmt, nicht zu einer Änderung der Variable H führen, die die Heilung der Krankheit des Patienten repräsentiert.

Schon diese knappe Überlegung zeigt, dass der Schluss von einer These der kausalen Geschlossenheit, der zufolge jedes physische Ereignis eine hinreichende physische Ursache hat, auf eine interventionistische Kausalaussage, der zufolge jede Variable, die ein physisches Ereignis repräsentiert, eine andere solche physische Variable als Ursache hat, nicht unproblematisch ist.

Zweitens hat die These der kausalen Geschlossenheit zunächst eine recht eingeschränkte Reichweite: Sie besagt, dass *physische*$_{eng}$ Ereignisse hinreichende *physische*$_{eng}$ Ursachen haben. Betroffen sind dementsprechend zunächst ausschließlich Eigenschaften und Ereignisse, die Gegenstand physikalischer Theorien sind. Auch Prämisse (IE3) kann also bloß physische$_{eng}$ Ereignisse betreffen, insofern sie

[216] Vgl. hierzu Abschnitt 4.1.2.

[217] Woodward (2015: 313), vgl. für ähnliche Anmerkungen Woodward (2008a: 245–247, 252) und Raatikainen (2010: 360).

durch die These der kausalen Geschlossenheit motiviert wird. Dies schränkt die Reichweite des interventionistischen Exklusionsarguments bedeutend ein.[218] Ob die Reichweite der These der kausalen Geschlossenheit ausgeweitet werden kann, hängt davon ab, ob sich ein Prinzip der Aufwärtsverursachung verteidigen lässt. Auf diese Frage gehe ich in Abschnitt 4.4.10. noch ein.

Drittens setzt eine Folgerung von der These der kausalen Geschlossenheit auf eine interventionistische Kausalaussage voraus, dass es überhaupt interventionistische Kausalität im Bereich der Physik gibt. Gerade dies wird jedoch von einigen Autor*innen angezweifelt.[219]

Anstatt sich auf die These der kausalen Geschlossenheit zu berufen, könnten sich Verteidiger*innen des interventionistischen Exklusionsarguments für die Begründung der Prämisse (IEA3) auf die interventionistischen Definitionen kausaler Begriff berufen: Es gibt mögliche Interventionen auf BM, die W verändern. Daher ist BM nach den Maßstäben des Interventionismus eine Ursache für W.

Das Problem mit dieser Begründung ist, dass man sich auf den ersten Blick mit gleichem Recht darauf berufen kann, dass es Interventionen auf M gibt, die W verändern und somit M nach den Maßstäben des Interventionismus eine Ursache für W ist. Natürlich würde Baumgartner nun einwenden, dass dies voraussetzt, dass es mögliche Interventionen auf M hinsichtlich W gibt und dass es nur dann mögliche Interventionen auf M hinsichtlich W gibt, wenn BM keine Ursache für W ist. Wenn die korrelative Beziehung zwischen M und W jedoch realisierungsinsensitiv ist – Änderungen von BM, die zwischen bm1 und bm2 wechseln, führen nicht zu Änderungen von W – ist diese Situation symmetrisch: Es gibt dann auch keine Interventionen auf BM, die W verändern, insofern M eine Ursache für W ist. Denn jede Variable, die BM von bm0 auf bm1 oder bm2 setzen kann und somit Interventionen auf BM erlauben würde, die W verändern, ist statistisch abhängig von M. Wenn M also eine Ursache für W ist, gibt es unter diesen Umständen keine Interventionen auf BM hinsichtlich W.

Die Frage, ob M oder BM eine Ursache für W ist, kann daher, soweit ich sehe, nicht aufgrund von Überlegungen entschieden werden, die sich aus dem (bis hierher spezifizierten) Interventionismus ergeben. M kann nur eine Ursache für W sein, wenn es Interventionen auf M gibt, die W verändern. Solche Interventionen kann es aber nur geben, wenn BM keine Ursache für W ist. BM kann nur eine Ursache

---

[218] Baumgartner (2010: 360) macht auch deutlich, dass sein Argument nur ‚micro-effects' betreffen soll. Wir haben jedoch gesehen, dass das interventionistische Exklusionsprinzip von dieser Einschränkung nicht betroffen ist.

[219] Vgl. für Diskussion Pearl (2000: 350), Hitchcock (2007) und Woodward (2007).

für W sein, wenn es Interventionen auf BM gibt, die W verändern. Solche Interventionen kann es aber nur geben, wenn M keine Ursache für W ist. Es ist aus dem Interventionismus alleine nicht entscheidbar, welche Veränderungen wir als Interventionen zählen sollten. Die Annahme, dass BM Ursache für W ist, setzt voraus, dass wir die mit Interventionen auf BM verbundenen Änderungen in M nicht als Intervention zählen, so dass M keine Ursache für W sein kann. Gleichzeitig setzt die Annahme, dass M Ursache für W ist, voraus, dass wir die mit Interventionen auf M verbundenen Änderungen in BM nicht als Interventionen zählen, so dass BM keine Ursache für W ist.

Wie Kim deutlich macht, hat die These der kausalen Geschlossenheit im einfachen Exklusionsargument die Funktion, die Symmetrie des Exklusionsprinzips zu durchbrechen, so dass das physische Ereignis dem mentalen Ereignis als Ursache vorgezogen werden kann.[220] Diese Funktion kann die These der kausalen Geschlossenheit im Rahmen des interventionistischen Exklusionsarguments nicht übernehmen. Grund hierfür ist, dass die These der kausalen Geschlossenheit keine direkten Konsequenzen für interventionistische Verursachung hat. Auch kann die Symmetrie des Exklusionsprinzips im interventionistischen Exklusionsargument nicht durch Überlegungen durchbrochen werden, die sich aus dem Interventionismus selbst ergeben. Der Interventionismus zwingt nicht dazu, Änderungen von BM, die auch M verändern, als Interventionen zu zählen, aber Änderungen von M, die auch BM verändern, nicht als Interventionen zu zählen.

Wie lässt sich also entscheiden, ob M oder BM eine Ursache für W ist? Es gibt hier im Prinzip zwei Möglichkeiten: Entweder, man schwächt den Interventionismus so ab, dass sowohl M als auch BM als Ursachen für W zählen können. Oder man verstärkt den Interventionismus um ein Kriterium, das eine Entscheidung für M und gegen BM [oder eine Entscheidung für BM und gegen M] herbeiführen kann. Die erste Option führt zu einer kompatibilistischen Konzeption mentaler Verursachung, während die zweite Option zu einer inkompatibilistischen Konzeption mentaler Verursachung führt. In den folgenden beiden Abschnitten gehe ich auf die beiden Optionen ein.

[220] Vgl. Abschnitt 3.3.1.

4.4.9 Modifizierter Interventionismus: Ablehnung des interventionistischen Exklusionsprinzips und der Status der drei Prinzipien der Verursachung im Stufenmodell der Realität

Das interventionistische Exklusionsprinzip kann vermieden werden, indem die Definition einer Interventionsvariablen modifiziert wird. Woodward schlägt die folgende Modifikation für die Definition einer Interventionsvariablen vor:

(IV*) Eine Variable *I* ist eine Interventionsvariable auf eine Variable *X* hinsichtlich einer Variable *Y* genau dann, wenn gilt:

(I1) *I* verursacht *X*.

(I2) *I* ist ein Umschalter für *X*. Das heißt, bestimmte Werte von *I* sind so beschaffen, dass gilt: Wenn *I* diese Werte annimmt, hängt *X* nicht von den Werten anderer Ursachen von *X* ab, sondern hängt stattdessen nur vom Wert von *I* ab.

(I3*) Jeder kausale Pfad von *I* zu *Y* geht durch *X* oder durch eine Variable Z, die in einer Supervenienzbeziehung zu X steht. Das heißt, *I* ist keine direkte Ursache von *Y* und ist keine Ursache von Ursachen von *Y*, die distinkt von *X* sind, außer jenen Ursachen von *Y*, die in den gerichteten Pfad von *I* über *X* zu *Y* eingebaut sind oder in einer Supervenienzbeziehung zu *X* stehen; also (a) jenen Ursachen von *Y*, die Wirkungen von *X* sind und (b) jenen Ursachen von *Y*, die auf einem kausalen Pfad zwischen *I* und *X* liegen und die keine von *X* unabhängigen Auswirkungen auf *Y* haben und (c) jenen Ursachen von *Y*, auf denen *X* superveniert oder die auf X supervenieren.

(I4*) *I* ist statistisch unabhängig von jeder Variablen *Z*, die *Y* verursacht, die auf einem kausalen Pfad liegt, der nicht durch *X* geht und die nicht in eine Supervenienzrelation zu X steht.[221]

(I3*) und (I4*) enthalten Ausnahmeklauseln für Variablen, die in einer Supervenienzrelation stehen: Laut (I3*) ist es nun zulässig, dass eine Interventionsvariable für X hinsichtlich Y auch eine Variable verursacht, die Y verursacht, solange X auf dieser Variablen superveniert. Laut (I4*) muss eine Interventionsvariable für X hinsichtlich Y nun nicht mehr von jeder Ursache von Y, die auf einem kausalen Pfad liegt, der nicht durch X geht, unabhängig sein, sondern nur von solchen Ursachen von Y, die nicht in einer Supervenienzbeziehung zu X stehen. Dass es keine Variablen gibt, die M verursachen, ohne auch BM zu verursachen und

---

[221] Vgl. Baumgartner (2013b: 13), Woodward (2015: 333–334), Woodward (2017: 262). Woodward erwähnt explizit nur Ausnahmeklauseln für subveniente Variablen. Tatsächlich müssen die Ausnahmeklauseln aber sowohl für subveniente als auch für supervenierende Variablen gelten, um die gewünschten Ergebnisse zu erhalten.

ohne von BM statistisch abhängig zu sein, spricht also nicht mehr dagegen, dass es Interventionsvariablen für M im Sinne von (IV*) gibt.

Setzt man diesen modifizierten Begriff einer Intervention voraus, stellt sich M als Ursache für W relativ zu {M, W} heraus: Eine (IV*)-Intervention auf M ist möglich und verändert W. Zudem stellt sich auch BM als Ursache für W relativ zu {BM, W} heraus: Eine (IV*)-Intervention auf BM verändert W ebenfalls. W ist also eine gemeinsame Wirkung von M und BM. Also ist das interventionistische Exklusionsprinzip dann falsch: Obgleich M auf BM superveniert, haben M und BM gemeinsame Wirkungen.

Die Ausnahmeklauseln für Supervenienzbasen in (IV*) mögen auf den ersten Blick wie ad hoc-Modifikationen aussehen. Woodward führt jedoch zwei Überlegungen an, die diesen Eindruck zumindest abschwächen:

Erstens sind solche Ausnahmeklauseln schon deshalb nötig, weil sich Supervenienzbeziehungen zwischen Variablen auch aus logischen oder begrifflichen Beziehungen zwischen Variablen ergeben können. Die schon in Abschnitt 4.3.7. diskutierten Variablenmengen {TS, T, N} und {LC, HC, GC} sind hier wiederum passende Beispiele. Ein allgemeines interventionistisches Exklusionsargument würde auch solche Variablen betreffen und zur Konsequenz haben, dass z. B. TS (die Schwellenwert-Variable für die Temperatur der Wasserprobe) keine Ursache für N sein kann, wenn T (die Variable für die genaue Temperatur) eine Ursache für N (das Garen der Nudeln) ist. Wieder sind die Probleme also nicht spezifisch für den nicht-reduktiven Physikalismus.

Zweitens, so Woodward, würde die Forderung, dass Interventionen unabhängig von Supervenienzbasen sein müssen, nicht gut zur wissenschaftlichen Praxis passen. In randomisierten kontrollierten Experimenten – die ja als Vorbild für Woodwards Bestimmung des Interventionsbegriffs dienen – wird jedenfalls keine Anstrengung unternommen, das Unmögliche zu erreichen und die experimentellen Interventionen z. B. auf die Einnahme von Medikamenten frei von statistischer Abhängigkeit von der Mikrostruktur der Medikamente zu halten.[222] Dies ist kein Mangel der entsprechenden Studien. Hintergrund ist hier, dass Supervenienzbasen nicht als gewöhnliche Störfaktoren agieren können. Wäre die Intervention auf die Einnahme des Medikaments abhängig vom Alter der Patient*innen – so dass das Medikament ausschließlich an junge Patient*innen ausgegeben wird und das Placebo ausschließlich an alte Patient*innen –, könnte es nicht ausgeschlossen werden, dass die beobachtete Korrelation zwischen Medikamenteneinnahme und Heilung auf das Alter der Patient*innen zurückzuführen

[222] Vgl. Woodward (2015: 337), Woodward (2017: 258–261).

ist: Bei jüngeren Patient*innen würde sich eine Heilung auch ohne Medikamenteneinnahme einstellen und bei älteren Patient*innen würde sich die Heilung auch bei Einnahme des Medikaments nicht einstellen. Eine Entsprechung findet sich im Falle der Abhängigkeit der Intervention auf die Einnahme des Medikament von der Mikrostruktur des Medikaments jedoch nicht: Es gibt einfach keine möglichen Situationen, in denen eine Substanz mit der Mikrostruktur des Medikaments ohne das Medikament eingenommen wird. Es kann daher auch nicht vorkommen, dass die Einnahme einer Substanz mit der Mikrostruktur des Medikaments auch ohne Einnahme des Medikaments zur Heilung führt.

Baumgartner reagiert wie folgt auf diese Begründung der Ausnahmeklauseln:

> While correlations of variables that are related in terms of reductive variants of supervenience are indeed likely not to give rise to data confounding, it is far from clear whether the same holds for variables related in terms of non-reductive supervenience.[223]

Unter reduktiven Varianten von Supervenienz versteht Baumgartner dabei u. a. Supervenienzbeziehungen zwischen Variablen, die sich aufgrund von logischen Zusammenhängen, aufgrund von Konstitution oder aufgrund von einer Beziehung zwischen Determinablen und Determinaten ergeben. Baumgartner meint, dass sich die Supervenienzbeziehung zwischen M und BM von solchen Fällen in entscheidender Hinsicht unterscheidet. Hier solle es sich schließlich um eine nicht-reduktive Variante von Supervenienz handeln. Er stimmt Woodward auch darin zu, dass es keine sinnvolle Forderung ist, Interventionen auf die Einnahme des Medikaments von der Mikrostruktur des Medikaments unabhängig zu halten. Denn die Beziehung zwischen Medikament und Mikrostruktur sei reduktiv. Jedoch besteht er darauf, dass sich diese Überlegung nicht auf den Fall mentaler Verursachung übertragen lässt.

Ausschlaggebend für Woodwards Argumentation sind jedoch die *modalen* Verbindungen zwischen subvenienter und supervenierender Variablen. Diese sorgen dafür, dass subveniente Variablen nicht als Störfaktoren in Betracht kommen. Folgt man dieser Argumentation, ist es daher schwierig, einen Keil zu schlagen zwischen solche Supervenienzbeziehungen, die sich aus logischen Beziehungen zwischen Variablen ergeben und solchen Supervenienzbeziehungen, die sich aus ontologischer Abhängigkeit im nicht-reduktiven Physikalismus ergeben. Im nicht-reduktiven Physikalismus unterscheidet sich der Zusammenhang zwischen den Variablen TS und T von dem Zusammenhang zwischen den Variablen M und BM

[223] Baumgartner (2013b: 18).

nicht in modaler Hinsicht.[224] Die zugehörigen Supervenienzbeziehungen haben dieselbe modale Kraft: metaphysische Notwendigkeit.[225] Dies scheint mir der wesentliche Grund dafür zu sein, dass in beiden Fällen die subvenienten Variablen nicht als Störfaktoren in Frage kommen. Die Annahme, dass die physischen Basen mentaler Eigenschaften trotz der genannten modalen Verknüpfungen als Störfaktoren in Frage kommen, bedarf daher zumindest weiterer Argumentation.

Woodwards Modifikation der Definition einer Interventionsvariablen scheint also gut zum Grundgedanken des Interventionismus zu passen, sich bei der Bestimmung kausaler Begriffe an der wissenschaftlichen Praxis zu orientieren. Auch führt sie – wie bereits geschildert – zum gewünschten Ergebnis mit Blick auf mentale Verursachung im nicht-reduktiven Physikalismus: M ist eine Ursache für W. Eine IV*-Intervention auf M verändert W. Zu was für einem Gesamtbild mentaler Verursachung führt sie aber? Dieser Frage möchte ich im Folgenden nachgehen, indem ich den Status der drei in Abschnitt 3.3.5. eingeführten Prinzipien der Verursachung im Stufenmodell der Realität unter Voraussetzung des IV*-Interventionismus diskutiere.

Betrachten wie zunächst das folgende Prinzip der basalen Verursachung:

> **Interventionistisches Prinzip der basalen Verursachung**: Wenn eine Variable U eine Ursache für eine Variable W ist und U nicht-kausal auf einer Variable B superveniert, dann ist auch B eine Ursache für W.

Zunächst eine Überlegung, die für dieses Prinzip spricht: IV*-Interventionen auf M führen aufgrund der Supervenienzbeziehung notwendigerweise auch zu Veränderungen von BM. Zudem ist es plausibel, dass diese Veränderungen in BM ebenfalls so zu verstehen sind, dass sie durch IV*-Interventionen herbeigeführt sind. Die Interventionsvariable I für M hinsichtlich W ist zugleich eine Interventionsvariable für BM hinsichtlich W. Wenn I nun einen Wert annimmt, der

---

[224] Vertritt man hiergegen die Auffassung, dass es logisch mögliche Welten gibt, die metaphysisch unmöglich sind, stellt sich noch immer die Frage, warum eine metaphysisch notwendige Verknüpfung zwischen physischen Basen und mentalen Eigenschaften nicht hinreichen sollte, um physische Basen als Störfaktoren auszuschließen.

[225] Im (naturalistischen) Dualismus ist das natürlich anders. Hier wäre also ein möglicher Einstiegspunkt für ein Argument dafür, dass nur der nicht-reduktive Physikalismus sich gegen den Vorwurf des Epiphänomenalismus wehren kann, während der Dualismus von diesem Vorwurf in aller Stärke betroffen ist: Wenn im Dualismus die nomologische Verknüpfung zwischen mentalen Eigenschaften und ihren Basen mit Blick auf ihre modale Kraft mit der Kausalrelation vergleichbar ist, stellt sich die Frage, warum die Basen nicht – ebenso wie andere Ursachen – als Störfaktoren in Betracht kommen sollten.

verursacht, dass M den Wert m1 annimmt, wird dieser Wert von I zugleich verursachen, dass BM den Wert b1 oder b2 annimmt. Jede IV*-Intervention auf M ist also zugleich eine IV*-Intervention auf BM. Wenn es demzufolge eine IV*-Intervention auf M gibt, die W verändert, wird diese selbe Intervention zugleich eine IV*-Intervention auf BM sein, die W verändert. Also folgt daraus, dass M eine Ursache für W ist, auch, dass BM eine Ursache für W ist.

Nun eine Komplikation: Insofern BM Ursachen haben kann, die keine Ursachen von M sind – die hiermit verbundene Ablehnung des Prinzips der Aufwärtskausalität diskutiere ich gleich noch – folgt daraus, dass I eine IV*-Interventionsvariable für M ist, nicht, dass I eine IV*-Interventionsvariable für BM ist. Grund hierfür ist Bedingung (I2): I könnte ein Umschalter für M sein, ohne ein Umschalter für BM zu sein.

Wenn I zum Beispiel einen Wert i1 hat, der eindeutig festlegt, dass M den Wert m1 hat, folgt daraus noch nicht, dass I einen Wert hat, der eindeutig festlegt, ob BM den Wert bm1 oder den Wert bm2 hat. Wenn i1 der einzige Wert ist, der den Wert von M eindeutig (auf m1) festlegt, hat I dann keinen Wert, der einen Wert von BM eindeutig festlegt. Ob BM dann den Wert bm1 oder den Wert bm2 annimmt, könnte von anderen Ursachen von BM abhängen, deren Einfluss auf BM durch I nicht gebrochen wird. Also ist I dann ein Umschalter für M, aber kein Umschalter für BM.

Dieses Bedenken kann jedoch leicht ausgeräumt werden: Zwar ist nicht jede Interventionsvariable für M auch eine Interventionsvariable für BM. Jedoch lässt sich immer eine Variable konstruieren, die sowohl für M als auch für BM Interventionsvariable ist. Wenn I also z. B. drei Werte (so viele wie BM) hat, so dass jeder Wert von I eindeutig festlegt, welchen Wert BM annimmt, dann wird I notwendigerweise auch Werte haben, die die Werte von M eindeutig festlegen. Insofern I dann Umschalter für BM ist, ist I auch Umschalter für M. Und wenn Interventionen auf M, die in Werten einer so konstruierten Interventionsvariablen bestehen, zu Änderungen von W führen, werden dieselben Interventionen auch Interventionen auf BM sein und zu Änderungen von W führen.

Der IV*-Interventionismus bestätigt also das interventionistische Prinzip der basalen Verursachung. Hieraus folgt, dass mentale Variablen im IV*-Interventionismus niemals Wirkungen haben, die nicht zugleich Wirkungen physischer Variablen sind. Dies spricht gegen eine echte Autonomie mentaler Ereignisse.[226]

Es ist hinzuzufügen, dass der IV*-Interventionismus nicht die Konsequenz hat, dass M und BM die genau selben Wirkungen haben. Es kann Wirkungen von BM

[226] Vgl. Abschnitt 3.4.4.

geben, die keine Wirkungen von M sind. Der einfache Grund hierfür ist, dass es Variablen geben kann, die nur auf Änderungen der präzisen Realisierung von **M** reagieren. Sie ändern sich also durch Interventionen auf BM, die zwischen bm1 und bm2 wechseln, aber durch keine anderen Interventionen auf BM.

Betrachten wir als nächstes den Status eines interventionistischen Prinzips der Abwärtsverursachung:

> **Interventionistisches Prinzip der Abwärtsverursachung**: Wenn eine Variable U eine Ursache für eine Variable W ist und W nicht-kausal auf einer Variable B superveniert, dann ist U auch eine Ursache für B.

Auch dieses Prinzip wird durch den IV*-Interventionismus bestätigt: Jede Variable, die M verursacht, verursacht auch BM. Denn jede Änderung von M geht mit einer Änderung von BM einher. Wenn es daher eine Intervention auf eine Variable U gibt, die M verändert, wird diese Intervention auch BM verändern. U zählt dann sowohl als Ursache von M als auch als Ursache von BM.[227]

Auch hier ergibt sich eine ähnliche Komplikation wie schon bei der Begründung des Prinzips der basalen Verursachung: Auf den ersten Blick ist es möglich, dass es eine Intervention auf U hinsichtlich M gibt, die M und BM verändert, ohne dass es eine Intervention auf U *hinsichtlich BM* gibt, die M und BM verändert. Gegeben, dass BM Ursachen haben kann, die keine Ursachen von M sind, ist nicht jede Intervention auf U hinsichtlich M zugleich eine Intervention auf U hinsichtlich BM. Denn die Interventionsvariable für U könnte statistisch unabhängig von allen Ursachen von M sein, ohne statistisch unabhängig von allen Ursachen von BM zu sein.

Bei der richtigen Wahl der Interventionsvariablen lässt sich diese Komplikation jedoch wiederum ausräumen: Es wird immer eine Interventionsvariable auf U hinsichtlich BM geben, die auch eine Interventionsvariable auf U hinsichtlich M ist. Diese Interventionsvariable erlaubt dann Interventionen auf U, die M und BM verändern, so dass U sowohl für M als auch für BM Ursache ist.

Kommen wir zum Status des folgenden interventionistischen Prinzips der Aufwärtsverursachung:

> **Interventionistisches Prinzip der Aufwärtsverursachung**: Wenn eine Variable U eine Ursache für eine Variable W ist und eine Variable H nicht-kausal auf W superveniert, dann ist U auch eine Ursache für H.

[227] Vgl. auch Baumgartner (2013b: 19).

Das Prinzip ist nicht allgemein gültig, weil es eine Ursache U für BM geben kann, die nur Relevanz dafür hat, ob BM den Wert bm1 oder den Wert bm2 annimmt. U macht dann einen Unterschied dafür, wie **M** realisiert ist, aber nicht dafür, ob **M** überhaupt instantiiert ist.[228]

Jedoch wird jede Ursache U von BM, für die es Interventionen gibt, die BM von bm0 auf bm1 oder bm2 setzen oder von bm1 oder bm2 auf bm0 setzen, auch eine Ursache für M sein. Denn solche Interventionen auf U sind notwendig mit Änderungen von M verbunden. Es gibt dann also IV*-Interventionen auf U, die M verändern und U zählt als Ursache für M. Ein entsprechend eingeschränktes, interventionistisches Prinzip der Aufwärtsverursachung wird im IV*-Interventionismus also bestätigt.

Insgesamt führt der IV*-Interventionismus also zu einer weitreichenden Vervielfältigung von Verursachungsbeziehungen: Jede Wirkung einer mentalen Variable hat (aufgrund des interventionistischen Prinzips der basalen Verursachung) auch jede subveniente physische Variable als Ursache. Jede Ursache einer mentalen Variable ist (aufgrund des interventionistischen Prinzips der Abwärtsverursachung) auch Ursache einer subvenienten physischen Variable. Und in vielen Fällen – jedoch nicht allgemein – zählen die Ursachen subvenienter physischer Variablen auch als Ursachen der zugehörigen supervenierenden mentalen Variablen.

Da das interventionistische Exklusionsprinzip im IV*-Interventionismus jedoch nicht gültig ist, führt diese Vervielfältigung von Kausalbeziehungen nicht zu erweiterten Exklusionsargumenten. Auch führt sie – wenn wir Woodwards Begründung dafür folgen, dass eine Unabhängigkeit von Interventionen von Supervenienzbasen nicht gerechtfertigt ist – nicht zu gefährlichen kausalen Fehlschlüssen.

Der IV*-Interventionismus ist jedoch mit keinerlei Diskriminierungen zwischen den zahlreichen, in Supervenienzbeziehungen stehenden Ursachen für eine Variable verbunden: Susis Armbewegung wird im selben Sinne von ihrem Wunsch, den neurologischen Prozessen in ihrem Hirn und den zugrundeliegenden fundamentalen mikrophysikalischen Vorgängen verursacht. Die mit der Idee der Proportionalität verbundene Intuition, dass Susis Wunsch als Ursache für ihre Armbewegung besser geeignet ist als die zugrundeliegenden mikrophysikalischen Vorgänge, wird im IV*-Interventionismus (bis hierher) also nicht eingefangen. Im

[228] Solche Variablen haben zudem Werte, die hinreichend dafür sind, dass M den Wert m1 annimmt. Jedoch sind sie keine Ursachen für M, weil eine Intervention auf diese Variablen nicht zu einer Änderung von M führt.

folgenden Abschnitt geht es nun um die Frage, welchen Platz solche Diskriminierungen im Rahmen einer interventionistischen Theorie mentaler Verursachung einnehmen können.

4.4.10 Proportionalität im Interventionismus

Sowohl der IV-Interventionismus als auch der IV*-Interventionismus führen zu Auffassungen von Verursachung, in denen Überlegungen zu Proportionalität keine Rolle spielen. Es stellt sich nun die Frage, ob die beiden Versionen des Interventionismus um Proportionalitätsbedingungen ergänzt werden können. Im Lichte der in Abschnitt 4.1. eingeführten Unterscheidung zwischen einer starken Proportionalitätsforderung, der zufolge Proportionalität eine notwendige Bedingung für Verursachung ist, und einer schwachen Proportionalitätsforderung, der zufolge Überlegungen zu Proportionalität einen Qualitätsunterschied zwischen verschiedenen Ursachen ziehen können, und den bisherigen Überlegungen zum interventionistischen Exklusionsargument, ergeben sich die folgenden zu diskutierenden Möglichkeiten:

Erstens könnte man den IV-Interventionismus um eine starke Proportionalitätsforderung ergänzen, um eine inkompatibilistische Erwiderung auf das interventionistische Exklusionsargument zu entwickeln.[229] Zweitens könnte man den IV*-Interventionismus um eine schwache Proportionalitätsforderung ergänzen, um den Eindruck abzuschwächen, dass er der Autonomie höherstufiger Eigenschaften nicht gerecht wird.[230] Ich argumentiere im Folgenden dafür, dass die erste Option nicht verteidigt werden kann. Die zweite Option ist hingegen vielversprechend.

---

[229] Eine schwache Proportionalitätsforderung kann diese Funktion nicht erfüllen. Das interventionistische Exklusionsprinzip etabliert schließlich, dass Variablen, die in einer Supervenienzbeziehung stehen, keine gemeinsamen Wirkungen haben. Um als Inkompatibilist*in dieses Prinzip akzeptieren zu können und an mentaler Verursachung festzuhalten, muss man argumentieren, dass die subveniente Variable BM *keine* Ursache für W ist, weil sie die Proportionalitätsforderung nicht erfüllt. Eine schwache Proportionalitätsforderung könnte jedoch allenfalls etablieren, dass BM keine *gute* Ursache für W ist. Ich diskutiere die Option, den IV-Interventionismus um eine schwache Proportionalitätsbedingung zu ergänzen, daher nicht weiter.

[230] Die Option, den IV*-Interventionismus um eine starke Proportionalitätsforderung zu ergänzen, diskutiere ich nicht, weil sie mit denselben Problemen zu kämpfen hätte, die schon dafür sorgen, dass der IV-Interventionismus sich nicht auf eine starke Proportionalitätsforderung berufen kann.

Die Diskussion des interventionistischen Exklusionsprinzips und des darauf aufbauenden Exklusionsarguments in den Abschnitten 4.4.7. und 4.4.8. hat gezeigt, dass der IV-Interventionismus ein interventionistisches Exklusionsprinzip bestätigt, jedoch keinerlei Mittel bereitstellt, zu entscheiden, welche vermeintliche Ursache hiervon betroffen ist: M und BM sehen beide wie legitime Ursachen für W aus. Gegeben das interventionistische Exklusionsprinzip können jedoch nicht sowohl M als auch BM Ursachen für W sein. Da schon die Frage, ob Manipulationen von M bzw. BM als Interventionen zählen können, davon abhängt, ob M bzw. BM Ursachen für W sind – also der interventionistische Test für jede Ursache von einem vorhergehenden Urteil über die andere Ursache abhängt – scheint eine Entscheidung innerhalb des IV-Interventionismus nicht möglich.

Verstärkungen des IV-Interventionismus könnten jedoch eine eindeutige Entscheidung herbeiführen. Insbesondere könnte eine starke Proportionalitätsforderung – ähnlich zu der von List und Menzies[231] – dazu führen, dass M in vielen Fällen als Ursache für W bevorzugt wird, während BM als Ursache für W ausgeschlossen wird. Dies legt also die folgende Verstärkung des IV-Interventionismus nahe:

> **IV-Interventionismus + Proportionalität (IVIP)**: Eine Variable X ist eine Ursache für eine Variable Y genau dann, wenn gilt: (i) es gibt eine Variablenmenge V, relativ zu der X eine beitragende Ursache für Y im Sinne des IV-Interventionismus ist und (ii) X ist proportional für Y.

Wenn BM nicht proportional für W ist, wird BM dann als Ursache für W ausgeschlossen. Entsprechend verlangt die Definition einer Interventionsvariablen (IV) auch nicht, dass Interventionsvariablen auf M unabhängig von BM sind: Interventionsvariablen auf M hinsichtlich W müssen nur unabhängig von pfadunabhängigen *Ursachen* für W sein. Da BM aber (aufgrund fehlender Proportionalität) keine Ursache für W ist, müssen Interventionsvariablen auf M hinsichtlich W nicht unabhängig von BM sein. Also gibt es Interventionsvariablen und Interventionen auf M hinsichtlich W. Da wir davon ausgehen, dass diese Interventionen auf M W verändern, steht der Annahme, dass M W verursacht, nun nichts mehr im Wege. Wenn M dann also zusätzlich auch proportional zu W ist, ist M auch Ursache für W.

Um diese Argumentation zu plausibilisieren, muss mehr darüber gesagt werden, wie Proportionalität innerhalb des Interventionismus verstanden werden kann. Was kann also damit gemeint sein, dass M proportional für W ist, während BM nicht proportional für W ist?

---

[231] Vgl. meine Diskussion in den Abschnitten 4.3.12. bis 4.3.14.

Eine mögliche Antwort hierauf hat in der Diskussion in den Abschnitten 4.4.7 – 4.4.9. bereits eine Rolle gespielt: Die Beziehung zwischen M und W ist realisierungsinsensitiv: Solange M den Wert m0 annimmt, nimmt auch W den Wert w0 an. Und solange M den Wert m1 annimmt, nimmt auch W den Wert w1 an. Ob BM dabei den Wert bm1 oder bm2 annimmt, ist irrelevant für dieses Korrelationsmuster. Wie **M** also realisiert ist, spielt keine Rolle für die Verbindung zwischen M und W.

Ausgehend von dieser Beobachtung lässt sich eine Forderung der Proportionalität formulieren, die sich in ihrer Grundidee an der Proportionalitätsforderung von List und Menzies orientiert. Eine Variable X ist proportional für eine Variable Y, wenn die Korrelation zwischen X und Y realisierungsinsensitiv ist.

Woodward schreibt zum Begriff der Realisierungsinsensitivität das Folgende:

> What is required is the existence of a relationship that both involves a dependency between the upper level variables (different values of M1, produced by interventions map into different values of M2) and that is realization independent in the sense that it continues to stably hold for a range of different realizers of these values of M1 and M2. It is the presence of this sort of realization independent dependency relationship (henceforth RIDR) that ensures that interventions that change M1 are stably associated with changes in M2 – hence that M1 causes M2.

Lässt sich diese Idee der Realisierungsinsensitivität präzisieren? Man kann es mit der folgenden Definition versuchen:

> **Realisierungsinsensitive Korrelation 1**: Eine Korrelation zwischen einer Variable X und einer Variable Y ist realisierungsinsensitiv genau dann, wenn gilt: Für jede Variable Z, auf der X superveniert und die die Basen der in X repräsentierten Eigenschaften repräsentiert, gilt: Keine Intervention auf Z, die X unberührt lässt, führt zu einer Änderung von Y.[232]

Ein Problem mit dieser Definition: Es ist nach dieser Definition nicht nur die Korrelation zwischen M und W realisierungsinsensitiv, sondern scheinbar auch

[232] Vgl. Hoffmann-Kolss (2014: 58): „Realization-independent dependency relationship: There is a realization-independent dependency relationship between two variables X and Y belonging to some causal graph G […] iff (i) X is classified as causally relevant to Y according to the interventionist account and (ii) any intervention changing only the realizers of the properties represented by X, but not the value of X itself, while the values of all other variables in G are held constant, does not change the value or the probability distribution of Y." Ich beschränke mich hier auf Bedingung (ii), da der Kontext es erfordert, eine kausale Beziehung im Sinne des IV-Interventionismus nicht vorauszusetzen. Zudem verzichte ich zur einfacheren Diskussion auf die Festhalten-Klausel.

die Korrelation zwischen BM und W. Gehen wir zum Beispiel davon aus, dass BM auf einer Variable PM (siehe Tabelle 4.10) superveniert.

Tabelle 4.10 Variable PM

Variable	Werte	Interpretation der Werte
PM	pm0,	pm0 = Keine Basis für **BM1** oder **BM2** ist instantiiert.
	pm1,	pm1 = **PM1**, eine Basis von **BM1**, ist instantiiert.
	pm2,	pm2 = **PM2**, eine Basis von **BM1**, ist instantiiert.
	pm3,	pm3 = **PM3**, eine Basis von **BM2**, ist instantiiert.
	pm4	pm4 = **PM4**, eine Basis von **BM2**, ist instantiiert.

Die Korrelation zwischen BM und W – die darin besteht, dass Manipulationen auf BM, die zwischen bm0 und einem der Basiswerte bm1 oder bm2 wechseln, zu einer Änderung von W führen – ist von Interventionen auf PM, die BM unberührt lassen, nicht betroffen. Keine Intervention auf PM, die BM unberührt lässt (z. B. Interventionen, die zwischen pm1 und pm2 wechseln, oder Interventionen, die zwischen pm3 und pm4 wechseln), führt zu einer Änderung von W. Entsprechendes wird auch für andere Variablen gelten, auf denen BM superveniert. Also ist die Korrelation zwischen BM und W realisierungsinsensitiv.

Tatsächlich scheint die Realisierungsinsensitivität der Korrelation zwischen BM und W sogar aus der Realisierungsinsensitivität der Korrelation zwischen M und W zu folgen: Wenn M auf BM superveniert und BM auf PM superveniert, dann superveniert M auf PM. Variablensupervenienz ist transitiv. Wenn die Beziehung zwischen M und W realisierungsinsensitiv ist, gibt es also keine Interventionen auf PM, die M unberührt lassen und W verändern. Da die Menge der Interventionen auf PM, die BM unberührt lassen, eine Teilmenge der Interventionen auf PM sind, die M unberührt lassen, gibt es dann auch keine Interventionen auf PM, die BM unberührt lassen und W verändern.

Man beachte aber, dass es Interventionen auf PM gibt, die BM *nicht* unberührt lassen, und die ebenfalls nicht zu Änderungen in W führen. Dies ist mit Blick auf BM und M hingegen nicht der Fall. Es gibt keine Interventionen auf BM, die M ändern, und zu keiner Änderung von W führen. Dies kann man sich zunutze machen und Realisierungsinsensitivität wie folgt bestimmen:

> **Realisierungsinsensitive Korrelation 2**: Eine Korrelation zwischen einer Variable X und einer Variable Y ist realisierungsinsensitiv genau dann, wenn gilt: Für jede Variable Z, auf der X superveniert und die die Basen der in X repräsentierten Eigenschaften repräsentiert, gilt: (i) Keine Intervention auf Z, die X unberührt lässt, führt

zu einer Änderung von Y und (ii) Jede Intervention auf Z, die X *nicht* unberührt lässt, verändert auch Y.

Gegeben diese Bestimmung realisierungsinsensitiver Korrelation ist die Korrelation zwischen M und W realisierungsinsensitiv, während die Korrelation zwischen BM und W nicht realisierungsinsensitiv ist. Denn es gibt Interventionen auf PM – z. B. solche, die zwischen pm2 und pm3 wechseln – die BM nicht unberührt lassen und W dennoch nicht verändern.

Diese Definition hat die Konsequenz, dass eine Korrelation zwischen zwei Variablen X und Y nur realisierungsinsensitiv sein kann, wenn *jede* Veränderung von X mit einer Veränderung von Y korreliert ist. Würde M beispielsweise dreiwertig sein und neben der Instantiierung von **M** auch noch die Instantiierung von **M*** repräsentieren, wobei **M*** eine leichte Abwandlung von **M** ist und ebenfalls dazu führt, dass W den Wert w1 annimmt, wäre die Korrelation zwischen M und W nicht mehr realisierungsinsensitiv. Denn es gäbe dann Variablen, auf denen M superveniert und die sowohl verschiedene Basen von **M** als auch verschiedene Basen von **M*** repräsentieren. Wenn auf diese Variablen so interveniert wird, dass statt einer Basis von **M** eine Basis von **M*** instantiiert ist, zieht dies keine Änderungen von W nach sich. Diese Eigenschaft ist jedoch möglicherweise gerade wünschenswert: Variablen, die proportional für andere Variablen sind, sollten keine 'überflüssigen' Werte enthalten. Und ein Grund, warum BM nicht als proportional für W gelten sollte, scheint gerade zu sein, dass BM solche überflüssigen Werte enthält. Dieser Aspekt wird durch das Kriterium der Realisierungsinsensitivität eingefangen.

Kommen wir zurück zum IV-Interventionismus + Proportionalität. Wenn Proportionalität als Realisierungsinsensitivität verstanden wird, ergibt sich folgender Versuch:

IV-Interventionismus + Realisierungsinsensitivität (IVIR): Eine Variable X ist eine Ursache für eine Variable Y genau dann, wenn gilt: (i) es gibt eine Variablenmenge V, relativ zu der X eine beitragende Ursache für Y im Sinne des IV-Interventionismus ist und (ii) die Korrelation zwischen X und Y ist realisierungsinsensitiv.

Lässt sich auf dieser Grundlage eine inkompatibilistische Erwiderung auf das interventionistische Exklusionsargument entwickeln?

Die Idee wäre folgende: Wir haben BM und M als mögliche Ursachen für W, können aber nicht entscheiden, ob es mögliche Interventionen auf M hinsichtlich W oder stattdessen mögliche Interventionen auf BM hinsichtlich W gibt, die W verändern. Wir verlassen daher zunächst den IV-Interventionismus und stellen die

Frage, ob die Korrelation zwischen M und W bzw. die Korrelation zwischen BM und W realisierungsinsensitiv ist. Wir stellen fest, dass die Korrelation zwischen M und W, jedoch nicht die Korrelation zwischen BM und W realisierungsinsensitiv ist. Wir können daher ausschließen, dass BM eine Ursache für W ist, weil BM Bedingung (ii) in (IVIR) verletzt. Nun kehren wir mit dieser Einsicht ausgerüstet zurück zur Frage, ob es Interventionen auf M gibt, die W verändern oder stattdessen Interventionen auf BM, die W verändern. Da BM keine Ursache für W ist, müssen Interventionen auf M hinsichtlich W nicht statistisch unabhängig von BM sein. Dass jede Manipulation von M auch BM verändert, spricht also nicht dagegen, dass es Interventionen auf M hinsichtlich W gibt. Wir stellen fest, dass Interventionen auf M auch W verändern. Also ist M eine Ursache für W relativ zu {M, W} und daher auch simpliciter. Wir stellen nun wiederum die Frage, ob auch BM eine Ursache für W relativ zu {BM, W} ist – wenn wir von der Frage der Realisierungsinsensitivität einmal absehen und uns auf den ursprünglichen IV-Interventionismus konzentrieren. Unsere Antwort ist negativ: Jede Manipulation von BM, die W verändert, verändert auch M. Da M aber Ursache für W ist, müssen Interventionen auf BM statistisch unabhängig von M sein. Also gibt es keine Interventionen auf BM, die W verändern. Die relevanten Manipulationen von BM, die W verändern, zählen aufgrund ihrer statistischen Abhängigkeit von M nicht als Interventionen. Also ist BM keine Ursache für W relativ zu {BM, W} oder zu irgendeiner Erweiterung von {BM, W}. Wir sehen uns in unserem (anstoßgebenden) Urteil, dass BM keine Ursache für W ist, bestätigt. Tatsächlich erfüllt BM weder Bedingung (i) noch Bedingung (ii) in (IVIR).

Diese Argumentationsstrategie hat jedoch folgendes Problem: In bestimmten Strukturen gemeinsamer Verursachung, die Variablen involvieren, die in Supervenienzbeziehungen stehen, ist es nicht möglich, ausschließlich realisierungsinsensitive Verursachungsrelationen zuzulassen. Vera Hoffmann-Kolss gibt hierfür das folgende Beispiel:[233]

Ein Stück Schokolade beginnt zu schmelzen, sobald ein Wasserbad eine Temperatur von 35° C erreicht hat. Die Temperatur des Wassers kann nun auf zwei Weisen in einer Variable repräsentiert werden: Erstens kann man eine binäre Schwellenwert-Variable TS einführen, die den Wert ts0 einnimmt, wenn das Wasser eine Temperatur zwischen 1° C und 35° C hat, und den Wert ts1 annimmt, wenn das Wasser eine Temperatur zwischen 35° C und 100° C hat. Zweitens kann man eine mehrwertige Variable T einführen, die Werte annimmt, die der Temperatur des Wasser in ° C entsprechen. T hat also den Wert t1, wenn das

[233] Hoffmann-Kolss (2014: 59–60).

Wasser eine Temperatur von 1° C hat, den Wert t35, wenn das Wasser eine Temperatur von 35° C hat, und den Wert t100, wenn das Wasser eine Temperatur von 100° C hat. Die Variable TS superveniert offenbar auf der Variable T: Für jeden Wert von TS gibt es Werte von T, die sie necessitieren. Wenn T beispielsweise den Wert t100 hat, muss TS den Wert ts1 annehmen und wenn T beispielsweise den Wert t1 hat, muss TS den Wert ts0 annehmen.

Führen wir nun zunächst eine Variable S ein, die repräsentiert, ob das Stück Schokolade schmilzt (s0 = die Schokolade schmilzt nicht, s1 = die Schokolade schmilzt). Nach dem IV-Interventionismus + Realisierungsinsensitivität (IVIR) können wir T als Ursache für S ausschließen, weil die Korrelation zwischen T und S nicht realisierungsinsensitiv ist. Unter dieser Voraussetzung können wir TS als Ursache für S bezeichnen, weil die Korrelation zwischen TS uns S realisierungsinsensitiv ist und es Interventionen auf TS gibt, die S verändern. Soweit erhalten wir also das gewünschte Ergebnis.

Wenn wir nun aber eine andere Variable A betrachten, die die Anzeige eines Thermometers repräsentiert, entstehen Probleme. A nimmt Werte a1 – a100 an, je nach Anzeige eines in das Wasser getauchten Thermometers in ° C. Die Beziehung zwischen TS und A ist nun nicht realisierungsinsenitiv, so dass TS als Ursache für A ausgeschlossen werden kann. Die Beziehung zwischen T und A ist hingegen realisierungsinsensitiv. Das Problem ergibt sich aus der Frage, wie wir im Rahmen von (IVIR) mit Variablenmengen umgehen können, die sowohl S als auch A enthalten.

Die Variablen S und A stehen nicht in einer Supervenienzbeziehung und es gibt keine metaphysisch unmöglichen Kombinationen von Werten von S und A. Es ist metaphysisch möglich, dass das Thermometer eine Temperatur von 100° C anzeigt und die Schokolade dennoch nicht schmilzt. Das Kriterium der unabhängigen Fixierbarkeit spricht also nicht gegen die Zulässigkeit von Variablenmengen, die S und A enthalten.

Es gibt dennoch eine Korrelation zwischen S und A. Diese kann jedoch nicht durch eine direkte Verursachungsbeziehung zwischen S und A erklärt werden: Der Thermometerstand verursacht nicht, dass die Schokolade schmilzt. Vielmehr haben Thermometerstand und Schmelzen eine gemeinsame Ursache – nämlich die Temperatur des Wassers.[234]

[234] Dies wird auch durch die kausale Markov Bedingung bzw. durch das daraus folgende Prinzip der gemeinsamen Verursachung gestützt: Wenn zwischen zwei Variablen X und Y in einer Variablenmenge eine Korrelation besteht, ist entweder X Ursache für, Y Ursache für X, oder X und Y haben eine gemeinsame Ursache. Vgl. Hoffmann-Kolss (2014: 60), Hausman and Woodward (1999).

Es gibt nun, gegeben die Variablen T und TS, auf den ersten Blick zwei Möglichkeiten, diese gemeinsame Ursache für S und A in eine zulässige Variablenmenge einzuführen: Erstens kann man die Variablenmenge {T, S, A} betrachten. Zweitens kann man die Variablenmenge {TS, S, A} betrachten. Das Problem besteht jedoch darin, dass laut (IVIR) keine dieser beiden Variablenmengen eine gemeinsame Ursache für S und A einführt.

Relativ zu {T, S, A} ist T eine Ursache für A, jedoch keine Ursache für S. Denn die Korrelation zwischen T und S ist nicht realisierungsinsensitiv. Relativ zu {TS, S, A} ist TS eine Ursache für S, aber nicht für A. Denn die Korrelation zwischen S und A ist nicht realisierungsinsensitiv. Also wird relativ zu keiner der betrachteten Variablenmengen die Temperatur des Wassers als gemeinsame Ursache für das Schmelzen der Schokolade und den Thermometerstand repräsentiert. Dies ist kein gutes Ergebnis. Jede Variablenmenge, die A und S enthält, wird eine unerklärte und nicht auf gemeinsame Ursachen zurückzuführende Korrelation zwischen A und S enthalten.

Ein entsprechendes Problem wird jede Auffassung treffen, die eine starke Proportionalitätsforderung für interventionistische Verursachung postuliert. Wann immer wir zwei Variablen X und Y haben, die in einer Supervenienzbeziehung stehen, wird es Wirkungen Wx geben, die zu X proportional sind, und andere Wirkungen Wy, die zu Y proportional sind, wobei diese Wirkungen Wx und Wy zwar miteinander korreliert sind, jedoch nicht selbst in einer Supervenienzbeziehung zueinander stehen. Variablenmengen, die Wx und Wy enthalten, werden keine gemeinsamen Ursachen für Wx und Wy enthalten. Es gibt dann also eine Korrelation zwischen Wx und Wy, die weder durch gemeinsame Verursachung, noch durch Verursachungsrelationen zwischen Wx und Wy, noch auch durch nicht-kausale Supervenienz zu erklären sind.

Dies scheint mir tatsächlich ein guter Grund zu sein, eine starke Proportionalitätsforderung im Interventionismus zu vermeiden. Allerdings hat die Argumentation nach meiner Auffassung nicht dieselbe Kraft gegen eine schwache Proportionalitätsforderung, die auf dem IV*-Interventionismus aufbaut. Die schwache Proportionalitätsforderung diskriminiert zwischen verschiedenen Ursachen, anstatt bestimmte Variablen als Ursachen auszuschließen.[235] Sie hat daher nicht die Konsequenz, dass A und S keine gemeinsamen Ursache haben.

Im IV*-Interventionismus sind sowohl TS als auch T gemeinsame Ursachen für A und S: TS ist gemeinsame Ursache für A und S relativ zu {TS, A, S}. T

[235] Im Interventionismus fügt sie den kausalen Graphen und Modellen damit Informationen hinzu. Im Gegensatz dazu würde eine starke Proportionalitätsforderung ihren Ausdruck darin finden, dass sie die Konstruktion kausaler Graphen anleitet. Die kausalen Graphen würden dann bereits Informationen über Proportionalität enthalten.

ist gemeinsame Ursache für A und S relativ zu {T, A, S}. Die schwache Proportionalitätsforderung fügt dem hinzu, dass TS eine ‘bessere’ Ursache für S ist als T, weil die Beziehung zwischen TS und S realisierungsinsensitiv ist, während die Beziehung zwischen T und S nicht realisierungsinsensitiv ist. Zudem ist T eine ‘bessere’ Ursache für A, weil die Beziehung zwischen T und A realisierungsinsensitiv ist, nicht aber die Beziehung zwischen TS und A. Sowohl die Variablenmenge {TS, A, S} als auch die Variablenmenge {T, A, S} sind dann zulässige Repräsentationen der vorliegenden kausalen Struktur. Die beiden Variablenmengen haben jedoch unterschiedliche Vor- und Nachteile: {TS, A, S} stellt die Beziehung zwischen TS und A nicht ideal dar, aber dafür die Beziehung zwischen TS und S. {T, A, S} stellt die Beziehung zwischen T und S nicht ideal dar, dafür aber die Beziehung zwischen T und A. Man kann daher nicht absolut sagen, dass {TS, A, S} die bessere Repräsentation ist oder dass {T, A, S} die bessere Repräsentation ist. Dies hängt scheinbar auch davon ab, welche kausale Beziehung im gegebenen Kontext relevanter ist.[236] Man kann jedoch eindeutig die Stärken und Schwächen der beiden Variablenmengen identifizieren.[237]

Es ist zudem sinnvoll, den IV*-Interventionismus um eine schwache Proportionalitätsforderung zu ergänzen. Um die Autonomie mentaler Ursachen zu betonen, ist es wichtig, deutlich zu machen, in welchem Sinne es sich bei mentalen Ursachen in vielen Kontexten um ‘bessere’ Ursachen handelt als bei zugrundeliegenden physischen Ursachen.

4.4.11 Fazit

In diesem Unterkapitel habe ich untersucht, ob der Interventionismus zu einer zufriedenstellenden Konzeption mentaler Verursachung im nicht-reduktiven Physikalismus führt. Es hat sich dabei herausgestellt, dass der IV-Interventionismus – Woodwards ursprüngliche Formulierung der Theorie – ein interventionistisches Exklusionsprinzip bestätigt. Zudem hat der IV-Interventionismus keine Ressourcen, um zu entscheiden, welche Ursachen nach diesem Exklusionsprinzip vom

[236] Hoffmann-Kolss (2014: 61–62) macht eine entsprechende Anmerkung, die diese Herangehensweise nahelegt.

[237] Man beachte auch, dass die Unbestimmtheit oder Kontextabhängigkeit hier die Adäquatheit der Variablenmengen betrifft, nicht aber direkt die Beziehung der Realisierungsinsensitivität zwischen zwei feststehenden Variablen. Die Unbestimmtheit oder Kontextabhängigkeit ergibt sich aus der Frage, wie die Realisierungsinsensitivität verschiedener kausaler Zusammenhänge in einer Variablenmenge für die Bestimmung der Adäquatheit der Variablenmenge gewichtet werden sollte.

kausalen Ausschluss betroffen sind. Eine Auflösung dieser Unbestimmtheit durch eine starke Proportionalitätsforderung verbietet sich dabei, da sie zu Problemen im Zusammenhang mit bestimmten Strukturen gemeinsamer Verursachung führt.

Man kommt daher nicht umhin, die Definition einer Interventionsvariablen um eine Ausnahmeklausel für Variablen zu ergänzen, die in Supervenienzbeziehungen stehen. Das Ergebnis – der IV*-Interventionismus – führt zu einer recht liberalen Auffassung von Verursachung im nicht-reduktiven Physikalismus. Insbesondere bestätigt der IV*-Interventionismus ein interventionistisches Prinzip der basalen Verursachung: Mentale Variablen haben daher niemals Wirkungen, die nicht zugleich auch Wirkungen ihrer Supervenienzbasen sind. Zudem bestätigt der IV*-Interventionismus ein interventionistisches Prinzip der Abwärtsverursachung: Ursachen mentaler Variablen sind zugleich immer auch Ursachen ihrer Supervenienzbasen.

Diese weitreichende Vervielfältigung von Kausalrelationen im Interventionismus motiviert eine schwache Proportionalitätsforderung: Zwar haben die Wirkungen mentaler Ursachen im Interventionismus zahlreiche in Supervenienzbeziehung stehende Ursachen. Jedoch sind in den paradigmatischen Fällen die mentalen Ursachen die ‘besten’ oder ‘ausschlaggebenden’ Ursachen im Sinne der schwachen Proportionalitätsforderung. Mentalen Ursachen kommt daher durchaus eine (eingeschränkte) Autonomie zu.[238]

4.5 Kausale Abhängigkeit, kausale Produktion und der Vorwurf des Epiphänomenalismus

In den letzten beiden Unterkapiteln habe ich dafür argumentiert, dass die beiden vorherrschenden Abhängigkeitstheorien der Kausalität in ihrer Anwendung auf den nicht-reduktiven Physikalismus zur Konsequenz haben, dass es mentale Verursachung gibt. Die kontrafaktische Theorie der Kausalität und der Interventionismus führen daher zu einer Zurückweisung der ersten Prämisse des Vorwurfs des Epiphänomenalismus und damit zu einer kausalistischen Erwiderung (siehe Abschnitt 3.1.3.).: Wenn wir Kausalität also als kausale Abhängigkeit verstehen, hat der nicht-reduktive Physikalismus keinesfalls die Konsequenz, dass es keine mentale Verursachung gibt.

[238] Ich danke Vera Hoffmann-Kolss für hilfreiche Kommentare und Diskussionen zu diesem Unterkapitel.

Kritiker*innen des nicht-reduktiven Physikalismus können sich daran stoßen, dass die bis hierher verteidigte Erwiderung auf den Vorwurf des Epiphänomenalismus auf einem Verständnis von Kausalität als kausaler Abhängigkeit beruht. Die Erwiderung scheint nur dann erfolgreich zu sein, wenn Kausalität *tatsächlich* kausale Abhängigkeit ist. Wenn man Kausalität hingegen als *kausale Produktion* versteht, gehen die letzten beiden Unterkapitel gewissermaßen am Thema vorbei. Was dann eigentlich gezeigt werden müsste, ist, dass mentale Ereignisse ihre Wirkungen *hervorbringen* und *kausale Arbeit* leisten. Was aber bis hierher nur gezeigt wurde, ist, dass mentale Ereignisse unterschied-machende Faktoren für andere Ereignisse sind.

Dieser Linie von Kritik an der entwickelten Erwiderung möchte ich mich in diesem letzten Unterkapitel zuwenden. Ich gehe dabei wie folgt vor:

In Abschnitt 4.5.1. unterscheide ich zwei Versionen des Vorwurfs des Epiphänomenalismus und präzisiere darauf aufbauend die Kritik an der entwickelten Erwiderung auf den Vorwurf des Epiphänomenalismus. Abschnitt 4.5.2. geht auf eine verbreitete Reaktion auf den Vorwurf des Produktions-Epiphänomenalismus ein. In Abschnitt 4.5.3. verdeutliche ich, dass bereits der Nachweis der Existenz mentaler kausaler Abhängigkeit einiges leisten kann, was üblicherweise von mentaler Verursachung verlangt wird. In Abschnitt 4.5.4. weise ich das für den Vorwurf des Produktions-Epiphänomenalismus zentrale Argument zurück und argumentiere dafür, dass die Existenz mentaler Produktion nicht notwendig für die Existenz von Handlungen ist.

4.5.1 Zwei Versionen des Vorwurfs des Epiphänomenalismus

Anknüpfend an die in Abschnitt 4.1.1. eingeführte Unterscheidung zwischen kausaler Abhängigkeit und kausaler Produktion lassen sich zwei verschiedene Versionen des Vorwurfs des Epiphänomenalismus formulieren: Eine Version, die sich auf kausale Abhängigkeit bezieht und eine Version, die sich auf kausale Produktion bezieht.

Der Vorwurf des Abhängigkeits-Epiphänomenalismus:

> **(VEA-1)** Wenn der nicht-reduktive Physikalismus wahr ist, dann gibt es keine mentale kausale Abhängigkeit.
>
> **(VEA-2)** Es gibt mentale kausale Abhängigkeit.

(VEA-C) Also: Der nicht-reduktive Physikalismus ist falsch.

Der Vorwurf des Produktions-Epiphänomenalismus:

(VEP-1) Wenn der nicht-reduktive Physikalismus wahr ist, dann gibt es keine mentale Produktion.

(VEP-2) Es gibt mentale Produktion.

(VEP-C) Der nicht-reduktive Physikalismus ist falsch.

Die Argumentation in den Unterkapiteln 4.3. und 4.4. hat gezeigt, dass die erste Prämisse des Vorwurfs des Abhängigkeits-Epiphänomenalismus falsch ist. Aus dem nicht-reduktiven Physikalismus folgt keineswegs, dass es keine mentale kausale Abhängigkeit gibt. Ganz im Gegenteil lässt sich zeigen, dass die vorherrschenden Theorien kausaler Abhängigkeit dazu führen, dass es mentale kausale Abhängigkeit unter Voraussetzung des nicht-reduktiven Physikalismus gibt.

Wir haben in Abschnitt 4.1.1. aber bereits gesehen, dass kausale Abhängigkeit auch ohne kausale Produktion vorkommen kann. Die dort diskutierten Fälle von negativer Kausalität und von Doppelverhinderung zeigen, dass kausale Abhängigkeit nicht hinreichend für kausale Produktion ist. Auch wenn die kontrafaktische Theorie der Kausalität und der Interventionismus also dazu führen, dass es mentale kausale Abhängigkeit gibt, folgt hieraus keineswegs, dass es mentale kausale Produktion gibt. Die bisherigen Überlegungen führen also noch nicht zu einer Zurückweisung des Vorwurfs des Produktions-Epiphänomenalismus.

Wie sollten nicht-reduktive Physikalist*innen also den Vorwurf des Produktions-Epiphänomenalismus zurückweisen? Es gibt – wie schon bei der originalen Version des Vorwurfs des Epiphänomenalismus – zwei Strategien zur Zurückweisung des Vorwurfs:

Erstens könnte man Prämisse (VEP-1) zurückweisen und argumentieren, dass der nicht-reduktive Physikalismus mit der Existenz mentaler Produktion kompatibel ist. Dies entspricht einer kausalistischen Erwiderung auf den Vorwurf des Produktions-Epiphänomenalismus.

Zweitens könnte man Prämisse (VEP-2) zurückweisen und akzeptieren, dass es im nicht-reduktiven Physikalismus keine mentale Produktion gibt. Mentale Ereignisse bringen ihre Wirkungen dann nicht hervor und leisten keine kausale Arbeit. Dies entspricht einer epiphänomenalistischen Erwiderung auf den Vorwurf des Produktions-Epiphänomenalismus.

Die epiphänomenalistische Erwiderung kann wiederum – wie in Abschnitt 3.5.3. ausgeführt – zwei Gestalten annehmen: Sie kann in einer radikalen und einer konservativen Form vertreten werden. Der radikale Produktions-Epiphänomenalismus würde besagen, dass die Nicht-Existenz mentaler Produktion weitreichende Konsequenzen für unser Selbst- und Menschenbild hat, mit denen wir wohl oder übel zu leben lernen müssen. Der konservative Produktions-Epiphänomenalismus würde hingegen bestreiten, dass aus der Nicht-Existenz mentaler Produktion unangenehme Konsequenzen folgen.

Ich führe in den kommenden Abschnitten einige Überlegungen zugunsten einer epiphänomenalistischen Erwiderung auf den Vorwurf des Produktions-Epiphänomenalismus an. Der Knackpunkt ist dabei, dass ein *konservativer* Produktions-Epiphänomenalismus vielversprechend ist, wenn eine Relation kausaler Abhängigkeit als ‚Ersatzrelation' bereitsteht. Die Idee einer Verteidigung eines konservativen Epiphänomenalismus unter Verweis auf eine Ersatzrelation habe ich bereits in Abschnitt 3.5.4. kurz angeschnitten. Hier wird diese Strategie nun in Bezug auf den Produktions-Epiphänomenalismus ausgeführt: Mentale kausale Abhängigkeit kann auch in Abwesenheit von mentaler kausaler Produktion alles leisten, was üblicherweise von mentaler Verursachung (bzw. mentaler Produktion) verlangt wird. Eine Ablehnung mentaler Produktion hat daher keine radikalen Konsequenzen.

Bevor ich dies aber weiter ausführe, beschäftige ich mich im kommenden Abschnitt mit einer anderen, verbreiteten Reaktion auf den Produktions-Epiphänomenalismus.

4.5.2 Die Dunkelheit des Produktionsbegriffs und die Frage nach der genuinen Kausalbeziehung

Eine verbreitete Reaktion auf den Vorwurf des Produktions-Epiphänomenalismus besteht in einer Grundsatzkritik an der Idee kausaler Produktion. So schreiben beispielsweise Christian List und Peter Menzies:

> Our response is that the concept of generation or production requires clarification. […]
>
> [U]nless a better explication can be given of causation as production, this notion can hardly play a significant role in the debate about mental causation.[239]

[239] List und Menzies (2009: 489).

Zudem bemerken sie, dass der Begriff der Produktion gelegentlich mit dem Begriff der hinreichenden Verursachung gleichgesetzt wird. Aber:

> If this is the interpretation of production or generation, however, then this concept is not a genuinely causal concept at all. There are well-known counter-examples to this view of causation.[240]

Hierbei stützen sie sich dann wiederum auf die Kritik an hinreichender Verursachung, die ich in Abschnitt 4.1.2. kurz dargestellt habe: Insbesondere werden irrelevante Faktoren als Ursachen klassifiziert. Der Lärm wird als Ursache für das Wiedererscheinen des Mondes nach der Mondfinsternis klassifiziert. Barry Loewer weist in demselben Kontext zudem auf das Problem der Komplexität hinreichender Ursachen hin: „If one wants to speak of 'production', it seems that it is the whole state and the laws that produce subsequent states."[241]

Eine typische Reaktion auf den Vorwurf des Produktions-Epiphänomenalismus besteht also in der Behauptung, dass der Begriff der kausalen Produktion dunkel und/oder nicht genuin kausal sei. Daher könne er keine tragende Rolle in einem Epiphänomenalismus-Vorwurf gegen den nicht-reduktiven Physikalismus spielen. Eine naheliegende Erwiderung auf diese Sorge besteht in einer Klärung und Verteidigung des Begriffs kausaler Produktion. Jaegwon Kim, der den Vorwurf des Produktions-Epiphänomenalismus stark macht, gründet sein Verständnis von Produktion auf zwei Quellen:

Erstens beruft er sich auf Elizabeth Anscombes klassischen Aufsatz 'Causality and Determination'.[242] Anscombes Verständnis von Kausalität ist eng an die Idee einer Produktion von Wirkungen durch Ursachen geknüpft: „Effects derive from, arise out of, come of, their causes"[243]. Hierbei argumentiert Anscombe jedoch explizit gegen eine Gleichsetzung oder enge Verwandtschaft von hinreichenden Bedingungen und Kausalität.[244] Gehen wir also von einem Anscomb'schen Verständnis von kausaler Produktion aus, scheint zumindest die Argumentationslinie,

---

[240] List und Menzies (2009: 489). Vgl. auch List und Mezies (2017: 277–278).

[241] Loewer (2015: 55). Vgl auch Loewer (2007: 253).

[242] Vgl. Kim (2007: 235).

[243] Anscombe (1993: 92).

[244] Vgl. Anscombe (1993: 92): „If A comes from B, this does not imply that every A-like thing comes from some B-like thing or set-up or that every B-like thing or set-up has an A-like thing coming from it; or that given B, A had to come from it, or that given A, there had to be B for it to come from. Any of these may be true, but if any is, that will be an additional fact, not comprised in A's coming from B. If we take 'coming from' in the sense of travel, this is perfectly evident."

nach der Produktion kein genuin kausaler Begriff ist, weil er mit hinreichender Verursachung gleichgesetzt wird, keine Schlagkraft zu haben.

Mit Bezug auf das Anscomb'sche Verständnis von Produktion schreiben List und Menzies entsprechend: „Anscombe treats productive causation as primitive, and neither she nor Kim elucidates the concept“[245]. Dieser Begriff sei also zu unverständlich, um eine tragende Rolle in einem Epiphänomenalismus-Vorwurf zu spielen. Man kann hier aber erwidern, dass einige Begriffe nun einmal primitiv sind. Und es ist durchaus möglich, dass wir eine hinreichende Bekanntschaft mit einem primitiven Begriff kausaler Produktion haben, auch wenn wir ihn nicht durch andere Begriffe definieren oder reduktiv analysieren können.

Zweitens bezieht sich Kim zur Erläuterung seines Verständnisses von kausaler Produktion auf Phil Dowes Erhaltungsgrößentheorie der Kausalität. Kim schreibt:

> These causal processes all involve real connectedness between cause and effect, and the connection is constituted by phenomena such as energy flow and momentum transfer, an actual movement of some (conserved) physical quantity.[246]

Wie in Abschnitt 4.1.1. geschildert, versteht Dowe kausale Prozesse als Weltlinien von Objekten, die eine Erhaltungsgröße haben und kausale Interaktionen als Überschneidungen von Weltlinien von Objekten, die mit einer Übertragung einer Erhaltungsgröße einhergehen. Die Grundidee ist dann, dass Ursachen und Wirkungen durch so verstandene kausale Prozesse und kausale Interaktionen miteinander verbunden sind und dass sich entsprechend immer eine Erhaltungsgröße von der Ursache zur Wirkung verfolgen lässt.[247] Die Idee eines Produzierens oder Hervorbringens findet ihren Ausdruck dann in der Übertragung dieser Erhaltungsgröße.

Ein Dowe'sches Verständnis von kausaler Produktion ist also nicht primitiv und auch nicht in einem problematischen Maße dunkel. Auch hier ist Produktion nicht mit hinreichender Verursachung gleichzusetzen: Mein Stoß gegen die weiße Kugel beim Billard ist über kausale Prozesse und Interaktionen mit dem Einlochen der 11er Kugel verbunden, ohne dass der Stoß für sich genommen hinreichend für das Einlochen wäre. Der Einwand, dass der Produktionsbegriff zu dunkel ist oder deshalb kein genuin kausaler Begriff ist, weil er mit hinreichender Verursachung gleichzusetzen ist, scheint hier also nicht zu greifen.

[245] List und Menzies (2009: 489).

[246] Kim (2007: 236).

[247] Vgl. Dowe (2000).

Natürlich könnte man nun argumentieren, dass kausale Produktion in diesem Sinne aus anderen Gründen kein genuin kausaler Begriff ist. Auch gegen Dowes Produktionstheorie der Kausalität gibt es „well-known counter-examples“[248].[249] Ich denke jedoch, dass die Existenz von Gegenbeispielen und die Frage, welcher Begriff oder welche Begriffe ‘genuin kausal’ sind, ohnehin nur von bedingter Relevanz in der Dialektik des Vorwurfs des Epiphänomenalismus sind.

Der Grund hierfür ist folgender: Selbst wenn wie auch immer verstandene kausale Produktion kein genuin kausaler Begriff ist, kann der Vorwurf des Produktions-Epiphänomenalismus ein ernstzunehmendes Problem für den nicht-reduktiven Physikalismus darstellen. Wenn aus dem nicht-reduktiven Physikalismus folgt, dass es keine mentale Produktion gibt und sich aus der Nicht-Existenz mentaler Produktion weitreichende Konsequenzen für unser Selbst- und Menschenbild ergeben, ist dies auch dann ein Problem für den nicht-reduktiven Physikalismus, wenn Produktion kein genuin kausaler Begriff ist. Eine Grundsatzkritik am Begriff kausaler Produktion kann daher für sich genommen nicht den Vorwurf des Produktions-Epiphänomenalismus zurückweisen. Es müsste noch immer gezeigt werden, dass aus dem nicht-reduktiven Physikalismus nicht folgt, dass es keine mentale Produktion gibt, oder dass wir ohne radikale Konsequenzen auf die Annahme mentaler Produktion verzichten können. Der Verweis darauf, dass der Begriff kausaler Produktion dunkel oder nicht-kausal ist, lässt offen, welche dieser Strategien mit Blick auf den Vorwurf des Produktions-Epiphänomenalismus verfolgt wird.

Dieser Punkt geht in beide Richtungen: Auch der Nachweis, dass kausale Abhängigkeit kein genuin kausaler Begriff ist, würde nicht zeigen, dass kausale Abhängigkeit keine wichtige Arbeit für die Zurückweisung des Vorwurfs des Epiphänomenalismus leisten kann. Man kann einen Produktions-Epiphänomenalismus vertreten, auch wenn *Produktion* die einzige ‚echte‘ Kausalbeziehung ist.[250] Denn mit den (laut Autoren wie Dowe und Kim eben

[248] List und Menzies (2009: 235).

[249] Vgl. Hitchcock (2009) für eine Kritik von Dowes Definitionen kausaler Prozesse und kausaler Interaktionen und Schaffer (2000) und Schaffer (2004b) für eine Kritik an Definitionen von Kausalität, die auf der Idee einer physischen Verknüpfung durch kausale Prozesse und kausale Interaktionen fußen.

[250] Hall (2004: 256) schreibt beispielsweise: „Production does seem, in some sense, to be the more ‘central’ notion“. Seine Einschätzung beruht allerdings auch auf einer Analyse kausaler Abhängigkeit durch eine einfache kontrafaktische Theorie, die Fälle von Präemption fehlklassifiziert.

nur Pseudo-kausalen) Relationen, die in der kontrafaktischen Theorie der Kausalität und der interventionistischen Theorie der Kausalität spezifiziert werden, steht eine hervorragende Ersatz-Relation für die Verteidigung eines *konservativen* Produktions-Epiphänomenalismus zur Verfügung. Der Produktions-Epiphänomenalismus hat keine radikalen Konsequenzen für unser Welt- und Menschenbild. Denn alles, was von einer ‚genuinen' – das hieße dann produktiven – Kausalrelation geleistet werden soll, kann auch von einer Beziehung der kausalen Abhängigkeit geleistet werden.

Es zeigt sich hier, dass die Frage nach der *korrekten* Theorie der Kausalität für die Zurückweisung des Vorwurfs des Epiphänomenalismus weniger bedeutsam ist, als häufig angenommen. Insbesondere ist es nicht so, als wäre die Plausibilität des nicht-reduktiven Physikalismus davon abhängig, welche Theorie der Kausalität sich in der unabhängigen Diskussion um die Metaphysik der Kausalität durchsetzt. Nicht-reduktive Physikalist*innen müssen nicht darauf hoffen, dass sich Abhängigkeitstheorien insgesamt als plausibler erweisen als Produktionstheorien. Vielmehr reicht es, zu zeigen, dass die Relation der kausalen Abhängigkeit – ganz gleich, ob es sich um eine genuin kausale Beziehung handelt – geeignet ist, um absurde Konsequenzen des nicht-reduktiven Physikalismus zu vermeiden.

Um dies weiter auszuführen, gehe ich im nächsten Abschnitt zunächst darauf ein, was der Nachweis mentaler kausaler Abhängigkeit bereits leisten kann – selbst wenn er nicht durch Produktionsbeziehungen untermauert ist.

4.5.3 Der Wert mentaler kausaler Abhängigkeit

Jonathan Schaffer vergleicht zwei Mechanismen, die einen Knopfdruck mit der Explosion einer Bombe verbinden könnten:

> Perhaps pressing the button generates an electrical current that connects to the bomb and triggers the explosion.
>
> Or perhaps pressing the button causes the absence of an inhibiting shield that had been preventing the source current from triggering the explosion.[251]

Im ersten Fall löst der Knopfdruck ein elektrisches Signal aus, das direkt zur Explosion der Bombe führt. In diesem Fall kann die Verbindung zwischen Knopfdruck und Explosion als eine Beziehung kausaler Produktion verstanden werden.

---

[251] Schaffer (2004b: 197–198). Vgl. auch Schaffer (2000: 285–287).

Im zweiten Fall hingegen ist der Knopfdruck mit der Explosion der Bombe über eine Doppelverhinderungsstruktur verbunden: Das Quellsignal ist auch ohne den Knopfdruck bereits vorhanden, wird jedoch nicht an die Bombe weitergeleitet, weil eine Abschirmung der Bombe dies verhindert. Der Knopfdruck entfernt dann die Abschirmung und verhindert so, dass die Abschirmung die Explosion der Bombe verhindert. In diesem Falle sind Knopfdruck und Explosion der Bombe nicht über eine Beziehung kausaler Produktion miteinander verbunden. Denn die Verbindung zwischen Knopfdruck und Explosion involviert ein negatives Ereignis, nämlich die Abwesenheit der Abschirmung.

Der Punkt, der anhand dieses Beispiels veranschaulicht werden kann, ist, dass der Unterschied zwischen den beiden Mechanismen in vielerlei Hinsicht irrelevant ist: Bei beiden Mechanismen macht der Knopfdruck *einen Unterschied* für die Explosion der Bombe. In beiden Fällen ist der Knopfdruck ein *effektives Mittel*, um die Explosion der Bombe hervorzurufen. In beiden Fällen kann man auf Grundlage der Information, dass der Knopf gedrückt wird, *vorhersagen*, dass die Bombe explodieren wird. Auch kann man in beiden Fällen eine Explosion der Bombe unter Bezug auf den Knopfdruck *erklären.* Schließlich kann man eine Person, die den Knopf drückt, auch in beiden Fällen *moralisch verantwortlich* für die Explosion der Bombe machen. Für all diese Dinge scheint es schlichtweg irrelevant zu sein, ob die Verbindung zwischen Knopfdruck und Explosion im Sinne einer kausalen Produktion verstanden werden kann, oder ob hier eine kausale Abhängigkeit *ohne Produktion* vorliegt.[252]

Diese Beobachtungen legen nahe, dass auch durch eine kausale Abhängigkeitsbeziehung zwischen *mentalen* und *behavioralen* Ereignissen bereits viel geleistet werden kann. Wir können hierdurch Erklärungen behavioraler Ereignisse durch mentale Ereignisse untermauern: Susis Wunsch, den Arm zu heben, erklärt, warum sie ihren Arm hebt. Wir können auf der Basis mentaler Ereignisse Vorhersagen über das Verhalten anderer Personen treffen: Wenn wir wissen, dass Susi den Wunsch hat, den Arm zu heben, ist dies ein guter Grund damit zu rechnen, dass sie den Arm heben wird. Und eine Einflussnahme auf die mentalen Ereignisse anderer Personen ist ein effektives Mittel, um ihr Verhalten zu ändern. All diese Punkte scheinen nicht darauf angewiesen zu sein, dass mentale Ereignisse ihre Wirkungen produzieren. Stattdessen reicht bereits das Bestehen einer kausalen Abhängigkeitsbeziehung für all dies hin.

Auch wenn das Bestehen einer kausalen Abhängigkeitsbeziehung zwischen mentalen Ereignissen und behavioralen Ereignisse also bereits einige radikale

[252] Vgl. Schaffer (2004b: 198–199).

Konsequenzen, die typischerweise mit dem Epiphänomenalismus in Zusammenhang gebracht werden (und teilweise für die Epiphobie verantwortlich sein mögen), abwenden kann, ist hiermit noch nicht gesagt, dass es nicht weitere Aspekte gibt, für die eine Produktionsbeziehung zwischen mentalen Ereignissen und behavioralen Ereignissen benötigt wird. In Abschnitt 3.5.2. habe ich das Argument aus der Existenz von Handlungen als das wichtigste Argument gegen den Epiphänomenalismus eingeführt. Im folgenden Abschnitt soll dieses Argument nun vor dem Hintergrund der Unterscheidung zwischen kausaler Produktion und kausaler Abhängigkeit betrachtet werden.

4.5.4 Kausale Abhängigkeit, kausale Produktion und die Existenz von Handlungen

Um Prämisse (VEP-2) des Vorwurfs des Produktions-Epiphänomenalismus (‘Es gibt mentale Produktion’) zu begründen, sollte man zeigen, dass die Ablehnung mentaler Produktion Konsequenzen hat, die man nicht akzeptieren kann. Die von Fodor geschilderte Epiphobie bezieht sich dann (auch) auf den Produktions-Epiphänomenalismus. Und der Produktions-Epiphänomenalismus käme nicht weniger als der Abhängigkeits-Epiphänomenalismus einem „Ende der Welt“ gleich.[253]

Wie in Abschnitt 3.5.2. geschildert, betrifft eine zentrale ‘Gefahr’ eines Epiphänomenalismus den Handlungsbegriff. Dies drückt sich in dem folgenden Argument aus der Existenz von Handlungen aus:

Das Argument aus der Existenz von Handlungen

(AH-1) Wenn es keine mentale Verursachung gibt, dann gibt es keine Handlungen.

(AH-2) Es gibt Handlungen.

(AH-3) Also: Es gibt mentale Verursachung.

Wenn nun gezeigt werden kann, dass dieses Argument aus der Existenz von Handlungen unter Voraussetzung eines Begriffs der kausalen Produktion funktioniert, ist dies sicherlich eines der stärksten möglichen Argumente für Prämisse

[253] Vgl. Fodor (1989: 77).

(VEP-2). Das Argument aus der Existenz von Handlungen gegen den Produktions-Epiphänomenalismus sieht wie folgt aus:

Das Argument aus der Existenz von Handlungen – Produktion

(AHP-1) Wenn es keine mentale Produktion gibt, dann gibt es keine Handlungen.

(AHP-2) Es gibt Handlungen.

(AHP-K) Also: Es gibt mentale Produktion.

Tatsächlich legt Jaegwon Kim seinem Vorwurf des Produktions-Epiphänomenalismus eben ein solches Argument zugrunde, wenn er schreibt:

> We care about mental causation because we care about human agency, and agency requires the productive/generative conception of causation. […] It seems to me that mere counterfactual dependence is not enough to sustain the causal relation involved in our idea of acting upon the natural course of events and bringing about changes so as to actualize what we desire and intend. An agent is someone who, on account of her beliefs, desires, emotions, intentions, and the like has the capacity to perform actions in the physical world: that is, to cause her limbs and other bodily parts to move in appropriate ways so as to bring about changes in the arrangement of objects and events around her – open a door, pick up the morning paper, and make a cup of coffee. It seems to me that without productive causation, which respects the locality/contiguity condition, such causal processes are not possible. These causal processes all involve real connectedness between cause and effect, and the connection is constituted by phenomena such as energy flow and momentum transfer, an actual movement of some conserved physical quantity.[254]

Die Idee ist hier also, dass wir nur dann handeln können, wenn wir unsere Handlungen *hervorbringen* oder *produzieren.* Da die Existenz mentaler kausaler Abhängigkeit jedoch nicht impliziert, dass wir unsere Handlungen hervorbringen oder produzieren, kann der Nachweis mentaler kausaler Abhängigkeit auch nicht die Existenz von Handlungen retten. Ohne mentale kausale Produktion gibt es keine Handlungen. Der Produktions-Epiphänomenalismus ist das Ende der Welt.

Ich sehe eine Reihe von Problemen mit dieser Argumentation, die ich im Folgenden schildern werde:

[254] Kim (2007: 236).

Erstens verweist Kim in dem obigen Zitat auf schwierig theoretisch einzuordnende Intuitionen über die Natur von Handlungen. Er schildert „our idea of acting upon the natural course of events“ und schreibt, dass „an agent is someone who […] has the capacity to […] cause her limbs […] to move“. Tatsächlich sind dies Ideen, die in unserem vortheoretischen Verständnis von Handlungen eine wichtige Rolle spielen. Sie theoretisch einzufangen, ist jedoch bekanntermaßen schwierig. Und es ist keinesfalls eindeutig, dass eine Produktionsbeziehung zwischen mentalen Ereignissen und Körperbewegungen hierbei von großer Hilfe wäre. Kims Formulierungen legen vielmehr nahe, dass erst eine *akteurskausalistische Theorie* von Handlungen die angesprochenen Intuitionen angemessen einfangen könnte: *Die Person* produziert ihre Handlungen und greift somit *von außen* in den natürlichen Verlauf der Welt ein. In *dieses* Paket von Intuitionen über Handlungen gehört sicherlich auch die Idee, dass die Person in einer *produktiven* Relation zu ihrer Handlung steht. Zugleich scheinen mir die Aussichten, diese Intuitionen im Rahmen einer naturalistisch ausgerichteten Handlungstheorie einzufangen, jedoch eher schlecht. Es ist daher fragwürdig, ob sich Kim auf solcherlei Intuitionen berufen kann, um ein reduktiv-physikalistisches Bild von Handlungen zu motivieren und es als vorteilhaft gegenüber einem nicht-reduktiv-physikalistischen Bild von Handlungen auszuweisen.

Naheliegend wäre es stattdessen, kausaler Produktion im Rahmen einer *kausalen Handlungstheorie* eine unverzichtbare Rolle zuzuschreiben. Zweitens sehe ich hierfür jedoch keine zwingenden Gründe. Die kausale Handlungstheorie kann ihre wesentlichen Vorzüge auch auf Grundlage einer Abhängigkeitstheorie der Kausalität etablieren. Dies kann durch die folgenden Überlegungen untermauert werden:

Die Unterscheidung zwischen Handlungen und bloßen Geschehnissen wird in der kausalen Handlungstheorie auf der Grundlage eines kausalen Kriteriums gezogen: Handlungen sind Ereignisse, die durch rationalisierende mentale Ereignisse *verursacht* werden. Dieses Unterscheidungskriterium zieht seine Attraktivität aus einer überzeugenden Anwendung auf paradigmatische Fälle: Susis Armbewegung wird als Handlung klassifiziert, weil sie von einem passenden Wunsch und einer passenden Überzeugung verursacht wird. Wenn Susi nießt, ist dies hingegen keine Handlung, weil dieses Ereignis nicht durch mentale Ereignisse verursacht wird, die es rational erscheinen lassen, zu nießen. Eine Deutung von Verursachung im Sinne kausaler Abhängigkeit in diesem Unterscheidungskriterium scheint gut geeignet zu sein, diesen Vorteil der kausalen Handlungstheorie zu erhalten: Susis Armbewegung ist kausal abhängig von rationalisierenden mentalen Ereignissen, aber ihr Nießen ist dies nicht. Dies kann schon auf der Grundlage einer einfachen kontrafaktischen Theorie der Kausalität etabliert werden: Wenn Susi nicht den Wunsch

gehabt hätte, den Arm zu heben, dann hätte sie nicht den Arm gehoben. Aber ihr Nießen ist nicht von rationalisierenden mentalen Ereignissen abhängig: Wenn Susi hätte nießen wollen hätte sie genießt und wenn sie *nicht* hätte nießen wollen, hätte sie ebenfalls genießt.

Eine klassische Motivation für die kausale Handlungstheorie ist zudem Donald Davidsons Beobachtung, dass „a person can have a reason for an action, and perform the action, and yet this reason not be the reason why he did it".[255] Al Mele gibt zur Veranschaulichung das folgende Beispiel:

> Two different things, T1 and T2, independently dispose Al to mow his lawn this morning. T2 has to do with schedule-related convenience and T2 with vengeance. Al wants to mow his lawn this week, and he believes that this morning is a convenient time, given his schedule for the week. But he also wants to repay his neighbour for the rude awakening he suffered recently when she turned on her mower at the crack of dawn, and he believes that mowing his lawn this morning would constitute suitable repayment. As it happens, Al's purpose in mowing his lawn this morning accords with one or the other of T1 and T2 but not both.[256]

Davidsons Herausforderung für alle Handlungstheorien besteht hier darin, eine Unterscheidung zwischen bloßen Gründen für eine Handlung und den Gründen, aus denen eine Handlung ausgeführt wurde, zu ziehen. Wir können zum Beispiel davon ausgehen, dass Al in Meles Beispiel den Rasen mäht, um seine Nachbarin zu verärgern, während seine Überlegungen zu seinem Wochenplan keine ausschlaggebende Rolle spielen. Die kausale Handlungstheorie kann diese Unterscheidung wiederum auf elegante Weise durch ein kausales Kriterium ziehen: Bloße Gründe für eine Handlung sind mentale Ereignisse, die eine Handlung zwar rationalisieren, aber nicht verursachen. Gründe, aus denen eine Handlung ausgeführt wird, sind hingegen mentale Ereignisse, die die Handlung sowohl rationalisieren als auch verursachen. Auch hier scheint eine Deutung von Verursachung im Sinne kausaler Abhängigkeit die gewünschte theoretische Rolle spielen zu können: Plausiblerweise sorgen bereits die einfachen kontrafaktischen Konditionale in Bezug auf das obige Beispiel für das gewünschte Ergebnis: Hätte Al nicht die Überlegungen zu seinem Wochenplan angestellt, hätte er den Rasen (aufgrund seines Wunsches nach Rache) noch immer gemäht. Hätte er hingegen den Wunsch nach Rache nicht gehabt, hätte er den Rasen nicht gemäht. Denn die Überlegungen zu seinem Zeitplan alleine hätten ihn nicht zur Handlung motiviert.

[255] Davidson (1963: 691).

[256] Mele (2013: 170).

Das einschlägigste Problem für die kausale Handlungstheorie besteht in Fällen abweichender Kausalketten. Gelegentlich verursachen rationalisierende mentale Ereignisse ein Ereignis, ohne dass dieses Ereignis plausibel als Handlung betrachtet werden kann. Das kausale Kriterium zur Abgrenzung von Handlungen ist daher nicht hinreichend. Davidson gibt hierfür das folgende Beispiel:

> A climber might want to rid himself of the weight and danger of holding another man on a rope, and he might know that by loosening his hold on the rope he could rid himself of the weight and danger. This belief and want might so unnerve him as to cause him to loosen his hold, and yet it might be the case that he never chose to loosen his hold, nor did he do it intentionally.[257]

Hier hat der Kletterer mentale Zustände, die einen Grund dafür konstituieren, das Seil loszulassen und den Mann in die Tiefe fallen zu lassen. Auch verursachen diese mentalen Zustände (über den Umweg der Nervosität), dass der Kletterer das Seil loslässt. Dennoch ist das Loslassen keine Handlung des Kletterers.

Ein starkes Argument für eine Ausdeutung der kausalen Handlungstheorie im Sinne kausaler Produktion würde sich sicherlich ergeben, wenn eine Produktionstheorie der Kausalität dieses Problem der abweichenden Kausalketten lösen könnte, während eine Abhängigkeitstheorie scheitert. Auch dies ist jedoch nicht zu erwarten. Denn die typische Reaktion auf Fälle kausaler Abweichung besteht nicht darin, die zugrundeliegende Kausalbehauptung zu bestreiten, sondern darin, das kausale Kriterium zur Abgrenzung von Handlungen um zusätzliche Bedingungen zu verstärken. Abweichende Kausalketten sind demnach *Kausalketten* – die eine wie auch immer spezifizierte zusätzliche Bedingung nicht erfüllen. Es ist daher nicht offenbar, dass unterschiedliche Kausaltheorien hier überhaupt etwas leisten können.

Wenn man aber den Weg wählt, zu bestreiten, dass die rationalisierenden mentalen Zustände im Fall des Kletterers verursachen, dass der Kletterer das Seil loslässt, scheinen Abhängigkeitstheorien der Kausalität sehr viel vielversprechender zu sein: Ein interessanter Ansatz beruht beispielsweise auf einer Anwendung einer Proportionalitätsforderung auf den Fall der abweichenden Kausalketten. So ließe sich, wie Neil McDonnell argumentiert, möglicherweise bestreiten, dass die rationalisierenden mentalen Ereignisse proportional zur Nervosität des Kletterers sind. Zumindest dies scheint aber eine Strategie zu sein, die nur im Rahmen einer Abhängigkeitstheorie zu verfolgen ist.[258]

---

[257] Davidson (1973: 79).

[258] Vgl. McDonnell (2015: 165). Es gibt jedoch Ansätze zur Lösung des Problems abweichender Kausalketten, die einer dispositionalen Theorie der Kausalität eine wichtige Rolle zuweisen. Vgl. Hyman (2014).

Eine abschließende Bewertung der Rolle der Unterscheidung zwischen kausaler Produktion und kausaler Abhängigkeit in der kausalen Handlungstheorie kann hier natürlich nicht geleistet werden. Die angeführten Überlegungen lassen jedoch keinen zwingenden Grund erkennen, der kausalen Handlungstheorie einen Begriff kausaler Produktion zugrunde zu legen. Stattdessen kann die kausale Handlungstheorie auf einer Abhängigkeitstheorie der Kausalität aufgebaut werden. Die kausale Handlungstheorie ist eine wichtige Motivation für Prämisse (AH-1) des ursprünglichen Arguments aus der Existenz von Handlungen. Wenn die kausale Handlungstheorie aber auf kausaler Abhängigkeit aufgebaut werden kann, überträgt sich diese Motivation nicht auf Prämisse (AHP-1) der Produktionsversion des Arguments aus der Existenz von Handlungen.

Ein drittes Problem mit Kims Plausibilisierung der Produktionsversion des Arguments aus der Existenz von Handlungen ist das folgende: Es gibt empirische Gründe, zu bestreiten, dass mentale Ereignisse ihre Wirkungen produzieren. Jonathan Schaffer schreibt folgendes über den tatsächlichen Mechanismus, der zu Muskelkontraktionen führt:

> [N]erve signals only cause muscle contractions […] by negative causation: the firing of the nerves causes a calcium cascade through the muscle fiber, which causes calcium-troponin binding, which causes the removal of tropomyosin from the binding sites on the actin which causes myosin-actin binding, and thereby causes the actin to be pulled in and the muscle to contract. […]
>
> Since all voluntary human actions are due to muscle contractions, it follows that all voluntary human actions (perhaps the most paradigmatic of all causes) involve negative causation.[259]

Laut Schaffer hat der Mechanismus, der zu Muskelkontraktionen führt, also eine Doppelverhinderungsstruktur: Denn ein wesentlicher Schritt in der Verursachung einer Muskelkontraktion ist die *Entfernung* von Tropomyosinfäden vom Aktinfilament. Ein Nervenimpuls von der motorischen Endplatte bewirkt zunächst die Ausschüttung von Calcium, was zu einer Entfernung der Tropomyosinfäden führt. Dies wiederum führt dazu, dass sich Myosin an das Aktin binden kann.[260] Der Nervenimpuls *verhindert* also, dass die Tropomyosinfäden *verhindern*, dass sich das Myosin an das Aktin bindet. Da es sich hier um einen Fall von Doppelverhinderung handelt, kann es sich nicht um eine Beziehung kausaler Produktion handeln.

[259] Schaffer (2004b: 200).

[260] Diese Theorie der Muskelkontraktion geht auf H. Huxley und Hanson (1954) und A.F. Huxley und Niedergerke (1954) zurück.

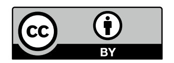
Da Handlungen üblicherweise mit Körperbewegungen einhergehen und Körperbewegungen Muskelkontraktionen involvieren, folgt aus dieser Beobachtung, dass Handlungen nicht durch neuronale Ereignisse produziert werden. Gerade auch in Versionen des *reduktiven* Physikalismus, in denen mentale Ereignisse mit neuronalen Ereignissen identifiziert werden, scheint hieraus zu folgen, dass mentale Ereignisse Handlungen nicht produzieren.[261]

Wenn Kim also damit recht hat, dass Handlungen produziert werden müssen und Produktion „an actual movement of some conserved physical quantity" erfordert, zeigt bereits eine genauere Betrachtung des Mechanismus, der zu Muskelkontraktion führt, dass es keine Handlungen gibt. Diese Konklusion ist dann *nicht* vom nicht-reduktiven Physikalismus abhängig, sondern ergibt sich gerade auch im reduktiven Physikalismus. Dies scheint ein guter Grund zu sein, die Forderung, dass Handlungen produziert werden müssen, aufzugeben.[262] Zudem ergibt sich hieraus ein guter Grund, eine kausale Handlungstheorie nicht auf den Begriff der kausalen Produktion aufzubauen: Denn insofern Handlungen tatsächlich *nicht* produziert werden, scheitert auch das auf kausaler Produktion basierende kausale Kriterium zur Abgrenzung von Handlungen: Susis Armheben ist ebensowenig von rationalisierenden mentalen Ereignissen produziert wie ihr Nießen.

Ich denke, dass die in diesem Abschnitt ausgeführten Überlegungen zeigen, dass die Produktionsversion des Arguments aus der Existenz von Handlungen scheitert: Aus der Nicht-Existenz mentaler Produktion folgt nicht, dass es keine Handlungen gibt. Hiermit bricht das zentrale Argument gegen einen Produktions-Epiphänomenalismus ein. Es sieht danach aus, dass ein *konservativer* Produktions-Epiphänomenalismus problemlos verteidigt werden kann: Die Existenz mentaler Produktion ist sicherlich kein Moore'scher Fakt und ich sehe nicht, warum die Nicht-Existenz mentaler Produktion das Ende der Welt sein sollte.

[261] Vgl. Russo (2015: 2192–1295).

[262] Vgl. auch Russo (2015) für weitere Diskussion der Konsequenzen für Kims Argument.


BY



5 Fazit

Ich fasse die Ergebnisse dieser Arbeit zusammen:

In Teil 2 habe ich mich mit dem Problem der Formulierung des nicht-reduktiven Physikalismus auseinandergesetzt. Dabei habe ich den nicht-reduktiven Physikalismus durch vier Kernthesen charakterisiert: *Erstens* sind unsere (Alltags-) Psychologie und die nicht-physikalischen Spezialwissenschaften ontologisch ernst zu nehmen. Dies drückt sich unter anderem in einer Festlegung auf die Existenz spärlicher mentaler und spezialwissenschaftlicher Eigenschaften und Ereignisse aus. *Zweitens* sind diese spärlichen mentalen und spezialwissenschaftlichen Eigenschaften und Ereignisse nicht mit physischen$_{\text{eng}}$ Eigenschaften und Ereignissen identisch. Dies liegt daran, dass sie *drittens* auf der Ebene von physischen$_{\text{eng}}$ Eigenschaften und Ereignissen multipel realisiert sind. *Viertens* schließlich sind mentale Eigenschaften und Ereignisse von physischen$_{\text{eng}}$ Eigenschaften und Ereignissen ontologisch abhängig. Diese Charakterisierung hat gezeigt, dass der nicht-reduktive Physikalismus eine klar abgegrenzte und zumindest prima facie attraktive Position ist.

Teil 3 hat sich darauf aufbauend mit dem Vorwurf des Epiphänomenalismus gegen den nicht-reduktiven Physikalismus beschäftigt. Hier habe ich zunächst eine einfache Rekonstruktion dieses Vorwurfs eingeführt: Die erste Prämisse des Vorwurfs des Epiphänomenalismus besagt dabei, dass der nicht-reduktive Physikalismus unweigerlich in einen Epiphänomenalismus führe, demzufolge es keine mentale Verursachung gibt. Die zweite Prämisse wendet sich gegen den Epiphänomenalismus. Dieser sei aufgrund seiner inakzeptablen Konsequenzen für unser Welt- und Menschenbild um jeden Preis zu vermeiden. Er würde – in Jerry Fodors Worten – einem Ende der Welt gleichkommen.

Zur Begründung der ersten Prämisse des Vorwurfs des Epiphänomenalismus sind Exklusionsargumente ausschlaggebend. Hierbei hat sich herausgestellt, dass das einfache Exklusionsargument eine stark eingeschränkte Reichweite

M. Rolffs, *Kausalität und mentale Verursachung*, Abhandlungen zur Philosophie,
https://doi.org/10.1007/978-3-662-66778-1_5

hat: Es kann bestenfalls nachweisen, dass mentale und spezialwissenschaftliche Ereignisse im nicht-reduktiven Physikalismus keine *im engen Sinne physischen* Wirkungen haben, nicht aber, dass sie keinerlei Wirkungen haben. Daher habe ich drei mögliche Erweiterungen des einfachen Exklusionsarguments ausgeführt, die auf drei Prinzipien der Verursachung im Stufenmodell der Realität beruhen. Für die Plausibilität der ersten Prämisse des Vorwurfs des Epiphänomenalismus sind die drei Prinzipien der Verursachung im Stufenmodell der Realität dabei nicht weniger wichtig als die Exklusionsprinzipien.

Mit Blick auf die zweite Prämisse des Vorwurfs des Epiphänomenalismus habe ich ein Argument aus der Existenz von Handlungen in den Mittelpunkt gestellt: Wenn es keine mentale Verursachung gibt, gibt es diesem Argument zufolge auch keine Handlungen. Diese Konsequenz des Epiphänomenalismus ist aber sicherlich inakzeptabel. Sie würde Fodors Einschätzung zur Bedeutsamkeit mentaler Verursachung gut begründen.

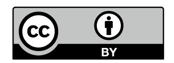
In Teil 4 der Arbeit habe ich mich schließlich ausführlich der Zurückweisung des Vorwurfs des Epiphänomenalismus gewidmet. Dabei stand die Idee von Kausalität als kausaler Abhängigkeit im Mittelpunkt: Ursachen werden hierbei als Faktoren aufgefasst, die einen Unterschied für das Eintreten ihrer Wirkungen machen. Diese Idee ist dabei von der Idee kausaler Produktion zu trennen. Denn Ursachen im Sinne kausaler Abhängigkeit müssen ihre Wirkungen nicht *hervorbringen* und keine *kausale Arbeit* leisten.

In den zentralen Unterkapiteln 4.3. und 4.4. habe ich dafür argumentiert, dass die beiden verbreitetsten Theorien kausaler Abhängigkeit dazu führen, dass es mentale kausale Abhängigkeit im nicht-reduktiven Physikalismus gibt. Dabei führen sowohl die kontrafaktische als auch die interventionistische Theorie zu einer recht liberalen Auffassung mentaler Verursachung, da beide Theorien auch die drei Prinzipien der Verursachung im Stufenmodell der Realität zumindest teilweise oder mit Einschränkungen bestätigen. Mentale Ereignisse verursachen somit nicht nur behaviorale und andere mentale Ereignisse, sondern auch $\text{physische}_{\text{eng}}$ Ereignisse. Zudem haben ihre Wirkungen neben den mentalen Ursachen auch $\text{physische}_{\text{eng}}$ Ursachen. Beide Theorien kausaler Abhängigkeit können aber durch eine schwache Proportionalitätsforderung ergänzt werden: Auf dieser Grundlage können innerhalb der zahlreichen in ontologischen Abhängigkeitsbeziehungen stehenden Ursachen wichtige Unterscheidungen getroffen werden. Mentale Ereignisse stellen sich dabei in paradigmatischen Fällen mentaler Verursachung gewissermaßen als die ‚besseren' Ursachen im Vergleich mit ihren $\text{physischen}_{\text{eng}}$ Basen heraus. In diesem Sinne kommt ihnen durchaus eine gewisse kausale Autonomie gegenüber ihren $\text{physischen}_{\text{eng}}$ Basen zu.

In Unterkapitel 4.5. schließlich habe ich die Unterscheidung zwischen kausaler Produktion und kausaler Abhängigkeit auf den Vorwurf des Epiphänomenalismus angewendet. Dabei hat sich gezeigt, dass sich auf der Grundlage des Ergebnisses, dass es mentale kausale Abhängigkeit gibt, die absurden Konsequenzen eines Epiphänomenalismus vermeiden lassen: Insbesondere kann eine kausale Handlungstheorie allem Anschein nach auf einer Relation der kausalen Abhängigkeit aufgebaut werden. Es ist daher nicht zu sehen, wieso ein Produktions-Epiphänomenalismus um jeden Preis vermieden werden sollte. Wenn der nicht-reduktive Physikalismus also die Konsequenz hat, dass es keine mentale Produktion gibt, sollte dies keinesfalls zur Ablehnung des nicht-reduktiven Physikalismus führen.

Vor dem Hintergrund dieses Ergebnisses lässt sich noch einmal veranschaulichen, warum die von Fodor diagnostizierte Epiphobie – die Angst vor dem Epiphänomenalismus – unbegründet ist: Wenn wir Kausalität als kausale Abhängigkeit verstehen, gleicht die Epiphobie der Angst davor, dass einem der Himmel auf den Kopf fällt: Denn der Abhängigkeits-Epiphänomenalismus wäre sicherlich schwierig zu akzeptieren und würde unser Welt- und Menschenbild ins Wanken bringen. Jedoch ist der Abhängigkeits-Epiphänomenalismus keine real drohende Gefahr. Er wird nicht durch den nicht-reduktiven Physikalismus impliziert. Wenn wir Kausalität hingegen als kausale Produktion verstehen, gleicht die Epiphobie eher der Angst vor dem Sonnenaufgang: Denn der Produktions-Epiphänomenalismus mag – ebenso wie der Sonnenaufgang – für nicht-reduktive Physikalist*innen kaum zu vermeiden sein. Jedoch gibt es auch hier keinen Grund, sich zu fürchten. Denn der Produktions-Epiphänomenalismus hat nicht das Potential, unser Welt- und Menschenbild ins Wanken zu bringen

Der nicht-reduktive Physikalismus kann den Vorwurf des Epiphänomenalismus also zurückweisen. Er hat nicht die Konsequenz, dass es keine mentale kausale Abhängigkeit gibt. Ob er hingegen dazu führt, dass es keine mentale Produktion gibt, hat keine große Relevanz für seine Plausibilität. Denn selbst wenn es keine mentale Produktion gibt, ist das nicht das Ende der Welt.


BY

Literaturverzeichnis

Alexander, Samuel (1927): *Space, Time, and Deity. Vol. 2*. London: Macmillan

Anscombe, G. E. M. (1971): *Causality and Determinism.* Cambridge: Cambridge University Press.

Anscombe, G. E. M. (1993): „Causality and Determination“. In: E. Sosa und M. Tooley (Hgg.): *Causation.* Oxford: Oxford University Press. S. 88–104.

Armstrong, David M. (1978): *Universals and Scientific Realism: A Theory of Universals Vol. II*. Cambridge: Cambridge University Press.

Árnadóttir, Steinvör und Crane, Tim (2013): „There is No Exclusion Problem“. In: E. J. Lowe, S. C. Gibb und R. D. Ingthorsson (Hgg.): *Mental Causation and Ontology*. Oxford: Oxford University Press. S. 248–266.

Audi, Paul (2013): „ How To Rule Out Disjunctive Properties“. In: *Noûs* 47 (4), S. 748–766.

Austin, John L. (1962): *Sense and Sensibilia.* Oxford: Oxford University Press.

Baker, Lynne Rudder (1988): „Cognitive suicide“. In: R. H. Grimm und D. D. Merrill (Hgg.): *Contents of Thought*. Chicago: University of Arizona Press. S. 401–413.

Baker, Lynne Rudder (1993): „Metaphysics and Mental Causation“. In: J. Heil und A. R. Mele (Hgg.): *Mental Causation.* Oxford: Oxford University Press. S. 75–96.

Baker, Lynne Rudder (2009): „Nonreductive materialism“. In: B. McLaughlin und A. Beckermann (Hgg.): *The Oxford Handbook of Philosophy of Mind*. Oxford: Oxford University Press. S. 109–127.

Bartels, Andreas (2007): „Wissenschaftlicher Realismus“. In: A. Bartels und M. Stöckler (Hgg.): *Wissenschaftstheorie – Ein Studienbuch.* Paderborn: mentis. S. 199–220.

Bartels, Andreas: „Modality in Physics and Metaphysics“. Unveröffentlichtes Manuskript.

Baumgartner, Michael (2009): „Interventionist Causal Exclusion and Non-reductive Physicalism“. In: *International Studies in the Philosophy of Science* 23 (2), S. 161–178.

Baumgartner, Michael (2010): „Interventionism and Epiphenomenalsim“. In: *Canadian Journal of Philosophy* 40 (3), S. 359–383.

Baumgartner, Michael (2013a): „A Regularity Theoretic Approach to Actual Causation“. In: *Erkenntnis* 78 (1), S. 85–109.

Baumgartner, Michael (2013b): „Rendering Interventionism and Non-Reductive Physicalism Compatible“. In: *Dialectica* 67 (1), S. 1–27.

Baysan, Umut (2018): „The Multiple Realization Book By Thomas W. Polger and Lawrence A. Shapiro“. In: *Analysis* 78 (1), S. 177–180.

M. Rolffs, *Kausalität und mentale Verursachung*, Abhandlungen zur Philosophie,
https://doi.org/10.1007/978-3-662-66778-1

Bechtel, William und Mundale, Jennifer (1999): „Multiple Realizability Revisited: Linking Cognitive and Neural States“. In: *Philosophy of Science* 66 (2), S. 175–207.
Beebee, Helen (2004): „Causing and Nothingness“. In: J. Collins, N. Hall und L. Paul (Hgg.): *Causation and Counterfactuals*. Cambridge, MA: MIT Press, S. 291–308.
Beebee, Helen (2006): *Hume on Causation*. London: Routledge.
Beebee, Helen, Hitchcock, Christopher und Price, Huw (Hgg.) (2017): *Making a Difference: Essays on the Philosophy of Causation*. Oxford: Oxford University Press.
Bennett, Jonathan (1987): „Event causation: The counterfactual analysis“. In: *Philosophical Perspectives* 1, S. 367–386.
Bennett, Jonathan (2003): *A Philosophical Guide to Conditionals*. Oxford: Oxford University Press.
Bennett, Karen (2003): „Why the exclusion problem seems intractable and how, just maybe, to tract it“. In: *Noûs* 37 (3), S. 471–497.
Bennett, Karen (2008): „Exclusion again“. In: J. Hohwy und J. Kallestrup (Hgg.): *Being Reduced: New Essays on Reduction, Explanation, and Causation*. Oxford: Oxford University Press. S. 280–307.
Bennett, Karen (2017): *Making Things Up*. Oxford: Oxford University Press.
Bernstein, Sara (2017): „Causal Proportions and Moral Responsibility“. In: D. Shoemaker (Hrsg.): *Oxford Studies in Agency and Responsibility, Volume 4*. Oxford: Oxford University Press. S. 165–182.
Berto, Francesco and Jago, Mark (2018): „Impossible Worlds“. In: E.N. Zalta (Hrsg.): *The Stanford Encyclopedia of Philosophy* (Fall 2018 Edition). URL = <https://plato.stanford.edu/archives/fall2018/entries/impossible-worlds/>.
Bickle, John (2020): „Multiple Realizability“. In: E. N. Zalta (Hrsg.): *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition). URL = <https://plato.stanford.edu/archives/sum2020/entries/multiple-realizability/>.
Birnbacher, Dieter und Hommen, David (2012): *Negative Kausalität*. Berlin: de Gruyter.
Blackburn, Simon W. (1984): „Supervenience revisited“. In: I. Hacking (Hrsg.): *Exercises in Analysis: Essays by Students of Casimir Lewy*. Cambridge: Cambridge University Press. S. 59–74.
Block, Ned (1980): „Introduction: What is Functionalism?“. In: Block, Ned (Hrsg.): *Readings in Philosophy of Psychology*. Cambridge: Harvard University Press. S. 171–184.
Block, Ned (2003): „Do causal powers drain away?“. In: *Philosophy and Phenomenological Research* 67 (1), S. 133–150.
Bontly, Thomas D. (2005): „Proportionality, causation, and exclusion“. In: *Philosophia* 32 (1–4), S. 331–348.
Bowie, G. Lee (1979): „The similarity approach to counterfactuals: Some problems“. In: *Noûs* 13 (4), S. 477–498.
Bryant, Amanda (2020): „Physicalism“. In: M.J. Raven (Hrsg.): *The Routledge Handbook of Metaphysical Grounding*. New York: Routledge. S. 484–500.
Bunzl, Martin (1979): „Causal overdetermination“. In: *Journal of Philosophy* 76 (3), S. 134–150.
Campbell, Neil (2008): *Mental Causation. A Nonreductive Approach*. New York: Peter Lang.
Carey, Brandon (2011): „Overdetermination And The Exclusion Problem“. In: *Australasian Journal of Philosophy* 89 (2), S. 251–262.

Carruthers, Peter (2016): „Higher Order Theories of Consciousness“. In: E. N. Zalta (Hrsg.): *The Stanford Encyclopedia of Philosophy* (Fall 2016 Edition). URL = <https://plato.stanford.edu/archives/fall2016/entries/consciousness-higher/>.

Chalmers, David (2017): „Naturalistic Dualism“. In: S. Schneider und M. Velmans (Hgg.): *The Blackwell Companion to Consciousness*. New York: John Wiley & Sons. S. 363–373.

Chalmers, David J. (1996): *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.

Chalmers, David J. (2003): „Consciousness and Its Place in Nature“. In: St. P. Stich und T. A. Warfield (Hgg.): *Blackwell Guide to the Philosophy of Mind*. Oxford, u.a.: Blackwell. S. 102–142.

Child, William (1993): „Anomalism, uncodifiability, and psychophysical relations“. In: *Philosophical Review* 102 (2), S. 215–245.

Chomsky, Noam (1968): *Language and Mind*. Cambridge University Press.

Chomsky, Noam (1980): *Rules and representations.* New York: Columbia University Press.

Christensen, Jonas und Baysan, Umut (2018): „Why Incompatibilism about Mental Causation is Incompatible with Non-reductive Physicalism“. In: *Inquiry: An Interdisciplinary Journal of Philosophy*, S. 1–23.

Churchland, Paul (1981): „Eliminative materialism and the propositional attitudes“. In: *Journal of Philosophy* 78, S. 67–90.

Churchland, Paul (1993): „Evaluating Our Self Conception“. In: *Mind and Language* 8 (2), S. 211–222.

Colyvan, Mark (1998) „Can the Eleatic Principle be Justified?“. In: *Canadian Journal of Philosophy* 28 (3), S. 313–335.

Correia, Fabrice (2013): „Metaphysical Grounds and Essence“. In: B. Schnieder, M. Hoeltje, und Steinberg, A. (Hgg.): *Varieties of Dependence: Ontological Dependence, Grounding, Supervenience, Response-Dependence (Basic Philosophical Concepts)*. München: Philosophia Verlag. S. 271 – 296.

Cowling, Sam (2017): „Resemblance“. In: *Philosophy Compass* 12 (4), e12401.

Crane, Tim (1991): „All God Has to Do“. In: *Analysis* 51 (4), S. 235–44.

Crane, Tim und Mellor, D.H. (1990): „There is no Question of Physicalism“. In: *Mind* 99 (394), S. 185–206.

Daly, Chris (2012): „Scepticism about Grounding“. In: F. Correia und B. Schnieder (Hgg.): *Metaphysical Grounding: Understanding the Structure of Reality*. Cambridge: Cambridge University Press. S. 81 – 100.

Davidson, Donald (1963): „Actions, Reasons, and Causes“. In: *Journal of Philosophy* 60 (23), S. 685–700.

Davidson, Donald (1969): „The individuation of events“. In: C.G. Hempel, D. Davidson, N. Rescher (Hgg.): *Essays in Honor of Carl G. Hempel*. Dordrecht: D. Reidel. S. 216–234.

Davidson, Donald (1970): „Mental events“. In: L. Foster und J. W. Swanson (Hgg.): *Experience and Theory*. Amherst: University of Massachusetts Press. S. 79–101. [Wiederabdruck in: B. Beakley und P. Ludlow: *The Philosophy of Mind. Classical Problems. Contemporary Issues.* Cambridge, MA: MIT Press. S. 137–149.]

Davidson, Donald (1973): „ Freedom to Act“. In: T. Honderich (Hrsg.): *Essays on Freedom of Action.* London: Routledge. [Wiederabdruck in: Davidson, Donald (1980): *Essays on Actions and Events*. Oxford: Oxford University Press. S. 63–82.]

Davidson, Donald (1985): „Reply to Quine on Events“. In: E. Lepore, B.P. McLaughlin (Hgg.): *Actions and Events: Perspectives on the Philosophy of Donald Davidson.* Oxford: Blackwell. S. 244–249.

Davidson, Donald (1993): „Thinking causes“. In: J. Heil und A. R. Mele (Hgg.): *Mental Causation.* Oxford: Oxford University Press. S. 3–18.

Dowe, Phil (2000): *Physical Causation.* Cambridge: Cambridge University Press.

Dowe, Phil (2009): „Causal Process Theories“. In: H. Beebee, C. Hitchcock und P. Menzies (Hgg.): *The Oxford Handbook of Causation.* Oxford: Oxford University Press. S. 213–233.

Dowell, Janice (2006): „Formulating the Thesis of Physicalism: An introduction“. In: *Philosophical Studies* 131 (1), S. 1–23.

Duncan, Michael, Miller, Kristie und Norton, James (2017): „Is Grounding a Hyperintensional Phenomenon?“. In: *Analytic Philosophy* 58 (4), S. 297–329.

Engelhardt, Jeff (2015): „What is the Exclusion Problem?“. In: *Pacific Philosophical Quarterly* 96 (2), S. 205–232.

Eronen, Markus I. (2012): „Pluralistic physicalism and the causal exclusion argument“. In: *European Journal for Philosophy of Science* 2 (2), S. 219–232.

Esfeld, Michael (2005): *Philosophie des Geistes. Eine Einführung*. Bern: Bern Studies in the History and Philosophy of Science.

Ewing, Alfred C. (1951): *The fundamental questions of philosophy.* New York: Routledge & Kegan Paul.

Fang, Wei (2018): „The case for multiple realization in biology“. In: *Biology and Philosophy* 33 (1–2), 3.

Ferey, Samuel und Dehez, Pierre (2016): *Overdetermined Causation Cases, Contribution and the Shapley Value.* Kent: Law Revue 91 (2), S. 637–658.

Fine, Kit (1994): „Ontological Dependence“. In: *Proceedings of the Aristotelian Society* 95 (1), S. 269–290.

Fodor, Jerry (1974): „Special Sciences, or Disunity of Science as a Working Hypothesis“. In: *Synthese* 28 (2), S. 97–115.

Fodor, Jerry A. (1989): „Making mind matter more“. In: *Philosophical Topics* 17 (11), S. 59–79.

Francescotti, Robert (2014): *Physicalism and the Mind*. Dordrecht: Springer.

Franklin-Hall, Laura R. (2016): „High-Level Explanation and the Interventionist's ‘Variables Problem’“. In: *British Journal for the Philosophy of Science* 67 (2), S. 553–577.

Frisch, Mathias (2014): *Causal Reasoning in Physics*. Cambridge: Cambridge University Press.

Funkhouser, Eric (2002): „Three varieties of causal overdetermination“. In: *Pacific Philosophical Quarterly* 83 (4), S. 335–351.

Funkhouser, Eric (2006): „The determinable-determinate relation“. In: *Noûs* 40 (3), S. 548–569.

Funkhouser, Eric (2007a): „Multiple Realizability“. In: *Philosophy Compass* 2 (2), S. 303–315.

Funkhouser, Eric (2007b): „A liberal conception of multiple realizability“. In: *Philosophical Studies* 132 (3), S. 467–494.

Garcia, Robert K. (2014): „Closing in on Causal Closure“. In: *Journal of Consciousness Studies* 21 (1–2), S. 96–109.

Gibb, Sophie (2012): „Nonreductive Physicalism and the Problem of Strong Closure“. In: *American Philosophical Quarterly* 49 (1), S. 29–42.
Gibb, Sophie (2015): „The Causal Closure Principle“. In: *Philosophical Quarterly* 65, S. 626–647.
Gibbons, John (2006): „Mental causation without downward causation“. In: *Philosophical Review* 115 (1), S. 79–103.
Gillett, Carl (2002): „The dimensions of realization: A critique of the standard view“. In: *Analysis* 62 (4), S. 316–323.
Gillett, Carl (2003): „The metaphysics of realization, multiple realizability, and the special sciences“. In: *Journal of Philosophy* 100 (11), S. 591–603.
Gozzano, Simone und Hill, Christopher S. (Hgg.) (2012): *New Perspectives on Type Identity: The Mental and the Physical*. Cambridge: Cambridge University Press.
Hall, Ned (2002): „Non-locality on the Cheap? A New Problem for Counterfactual Analyses of Causation“. In: *Noûs* 36 (2), S. 276–294.
Hall, Ned (2004): „Two concepts of causation“. In: J. Collins, N. Hall und L. Paul (Hgg.): *Causation and Counterfactuals*. Cambridge: MIT Press. S. 225–276.
Hall, Ned (2007): „Structural equations and causation“. In: *Philosophical Studies* 132 (1), S. 109–136.
Hannan, Barbara (1993): „Don’t Stop Believing: The Case Against Eliminative Materialism“. In: *Mind and Language* 8 (2), S. 165–178.
Harbecke, Jens (2008): *Mental causation. Investigating the Mind’s Powers in a Natural World*. Frankfurt a.M.: Ontos.
Harbecke, Jens (2013): „Mental Causation and the New Compatibilism“. In: *Abstracta* 7 (1), S. 55–71.
Harbecke, Jens (2014): „Counterfactual Causation and Mental Causation“. In: *Philosophia* 42 (2), S. 363–385.
Haug, Matthew C. (2011): „On the distinction between reductive and nonreductive physicalism“. *Metaphilosophy* 42 (4), S. 451–469.
Haug, Matthew C. (2019): „No microphysical causation? No problem: selective causal skepticism and the structure of completeness-based arguments for physicalism“. In: *Synthese* 196 (3), S. 1187–1208.
Hausman, Daniel M und Woodward, James (1999): „Independence, invariance and the causal Markov condition“. In: *British Journal for the Philosophy of Science* 50 (4), S. 521–583.
Heil, John (2003): *From an Ontological Point of View*. Oxford: Oxford University Press
Heil, John (2009): „Anomalous monism“. In: H. Dyke (Hrsg.): *From Truth to Reality: New Essays in Logic and Metaphysics*. London: Routledge. S. 85-98.
Hempel, Carl G. (1980): „Comments on Goodman’s „Ways of Worldmaking““. In: *Synthese* 45 (2), S. 193–199.
Hitchcock, Christopher (2001): „A tale of two effects“. In: *Philosophical Review* 110 (3), S. 361–396.
Hitchcock, Christopher (2007): „What Russell got right“. In H. Price und R. Corry (Hgg.): *Causation, Physics, and the Constitution of Reality: Russell’s Republic Revisited*. Oxford: Oxford University Press.
Hitchcock, Christopher (2009): „Problems for the Conserved Quantity Theory: Counterexamples, Circularity, and Redundancy“. In: *The Monist* 92 (1), S. 72–93.

Hitchcock, Christopher (2018): „Causal Models“. In: E.N. Zalta (Hrsg.): *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition). URL = <https://plato.stanford.edu/archives/sum2020/entries/causal-models/>

Hoffmann, Vera und Newen, Albert (2007): „Supervenience of Extrinsic Properties“. In: *Erkenntnis* 67 (2), S. 305–319.

Hoffmann-Kolss, Vera (2014): „Interventionism and Higher-level Causation“. In: *International Studies in the Philosophy of Science* 28 (1), S. 49–64.

Hoffmann-Kolss, Vera (2021): „Interventionism and Non-Causal Dependence Relations: New Work for a Theory of Supervenience“. In: *Australasian Journal of Philosophy*.

Hommen, David (2013): *Mentale Verursachung, innere Erfahrung und handelnde Personen. Eine Verteidigung des Epiphänomenalismus*. Paderborn: mentis.

Honderich, Ted (1982): „The Argument for Anomalous Monism”. In: *Analysis* 42, S. 59–64.

Horgan, Terence (1993): „From supervenience to superdupervenience: Meeting the demands of a material world“. In: *Mind* 102 (408), S. 555–586.

Horgan, Terence (2001): „Multiple reference, multiple realization, and the reduction of mind“. In: G. Preyer und F. Siebelt (Hgg.): *Reality and Humean Supervenience: Essays on the Philosophy of David Lewis*. Lanham u.a.: Rowman & Littlefield. S. 205–221.

Hume; David (1739/1888): *A Treatise of Human Nature. Reprinted from the original version in three volumes and edited, with an analytical index, by L.A. Selby Bigge*. Oxford: Oxford University Press.

Hüttemann, Andreas (2004): *What's Wrong With Microphysicalism*?. London: Routledge.

Hüttemann, Andreas (2013a). *Ursachen*. Berlin: de Gruyter.

Hüttemann, Andreas (2013b): „Einige Bemerkungen zum Prinzip der kausalen Abgeschlossenheit des Physischen“. In: J. Michel und G. Münster (Hgg.): *Die Suche nach dem Geist*. Paderborn: mentis. S. 35–54.

Hüttemann, Andreas (2020): „Processes, pre-emption and further problems“. In: *Synthese* 197 (4), S. 1487–1509.

Hüttemann, Andreas und Papineau, David (2005): „Physicalism decomposed“. In: *Analysis* 65 (1), S. 33–39.

Huxley, Andrew F. und Niedergerke, Rolf (1954): „Structural Changes in Muscle During Contraction: Interference Microscopy of Living Muscle Fibres“. In: *Nature* 173, S. 971–973.

Huxley, Hugh und Hanson, Jean (1954): „Changes in the Cross-Striations of Muscle during Contraction and Stretch and their Structural Interpretation”. In: *Nature* 173, S. 973–976.

Hyman, John (2014): „Desires, Dispositions and Deviant Causal Chains“. In: *Philosophy* 89 (1), S. 83–112.

Jackson, Frank (1982): „Epiphenomenal Qualia“. In: *The Philosophical Quarterly* 32, S. 127–136.

Jackson, Frank (1994): „Armchair metaphysics“. In: J. O'Leary-Hawthorne und M. Michael (Hgg.): *Philosophy in Mind*. Alphen aan den Rijn: Kluwer Academic Publishers. S. 23–42.

Jackson, Frank (1998): *From Metaphysics to Ethics. A Defence of Conceptual Analysis*. Oxford: Oxford University Press.

Jacob, Pierre (1997): *What Minds Can Do. Intentionality in a Non-Intentional World*. Cambridge: Cambridge University Press.

Kaiserman, Alex (2018): „Interventionism and Mental Surgery“. In: *Erkenntnis*, S.1–17.

Kim, Jaegwon (1976): „Events as property exemplifications“. In: M. Brand und D. Walton (Hgg.): *Action Theory*. Dordrecht: D. Reidel. S. 310 – 326. [Wiederabdruck in: Kim, Jaegwon (1993): *Supervenience and Mind: Selected Philosophical Essays*. Cambridge: Cambridge University Press. S. 33–52.]

Kim, Jaegwon (1984a): „Concepts of supervenience“. In: *Philosophy and Phenomenological Research* 45 (December), S. 153–176. [Wiederabdruck in: Kim, Jaegwon (1993): *Supervenience and Mind: Selected Philosophical Essays*. Cambridge: Cambridge University Press. S. 53–78.]

Kim, Jaegwon (1984b): „Epiphenomenal and supervenient causation“. In: *Midwest Studies in Philosophy* 9 (1), S. 257–70.

Kim, Jaegwon (1987): „ ‚Strong‘ and ‚global‘ supervenience revisited“. In: *Philosophy and Phenomenological Research* 48, S. 315–326.

Kim, Jaegwon (1989): „The myth of non-reductive materialism“. In: *Proceedings and Addresses of the American Philosophical Association* 63 (3), S. 31–47.

Kim, Jaegwon (1990): „Explanatory exclusion and the problem of mental causation“. In: E. Villanueva (Hrsg.): *Information, Semantics, and Epistemology*. Oxford: Blackwell. S. 36–56.

Kim, Jaegwon (1992): „Multiple Realization and the Metaphysics of Reduction”. In: *Philosophy and Phenomenological Research* 52 (1), S. 1–26.

Kim, Jaegwon (1993a): „Can supervenience and „non-strict laws“ save anomalous monism?“. In: J. Heil und A. R. Mele (Hgg.): *Mental Causation.* Oxford: Oxford University Press. S. 19–26.

Kim, Jaegwon (1993b): „The nonreductivist’s trouble with mental causation“. In: J. Heil und A. R. Mele (Hgg.): *Mental Causation.* Oxford: Oxford University Press. S. 189–210.

Kim, Jaegwon (1998): *Mind in a Physical World. An Essay on the Mind-Body Problem and Mental Causation.* Cambridge, MA: MIT Press.

Kim, Jaegwon (2002a): „The layered model: Metaphysical considerations“. In: *Philosophical Explorations* 5 (1), S. 2–20.

Kim, Jaegwon (2002b): „Responses“. In: *Philosophy and Phenomenological Research* 65 (3), S. 671–680.

Kim, Jaegwon (2005): *Physicalism, or Something Near Enough.* Princeton: Princeton University Press.

Kim, Jaegwon (2007): „Causation and mental causation“. In: B. P. McLaughlin und J. D. Cohen (Hgg.): *Contemporary Debates in Philosophy of Mind*. Oxford: Blackwell. S. 227–242.

Kim, Jaegwon (2011): „From Naturalism to Physicalism: Supervenience Redux“. In: *Proceedings and Addresses of the American Philosophical Association* 85 (2), S. 109–134.

Kment, Boris (2017): „Varieties of Modality“. In: E.N. Zalta (Hrsg.): *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition). URL = <https://plato.stanford.edu/archives/spr2017/entries/modality-varieties/>.

Koslicki, Kathrin (2015): „The Coarse-Grainedness of Grounding“. In: *Oxford Studies in Metaphysics* 9, S. 306–344.

Koslicki, Kathrin (2020): „Skeptical Doubts“. In: M.J. Raven (Hrsg.): *The Routledge Handbook of Metaphysical Grounding*. New York: Routledge. S. 164 – 179.

Krämer, Stephan und Schnieder, Benjamin (2017): „Grounding“. In: M. Schrenk (Hrsg.). In: *Handbuch Metaphysik*. Stuttgart: Metzler. S. 278–284.

Kripke, Saul A. (1980): *Naming and Necessity*. Harvard: Harvard University Press.

Kroedel, Thomas (2015a): „A simple argument for downward causation“. In: *Synthese* 192, S. 841–858.

Kroedel, Thomas (2015b): „Dualist Mental Causation and the Exclusion Problem“. In: *Noûs* 49 (2), S. 357–375.

Kroedel, Thomas (2019): *Mental Causation. A Counterfactual Theory*. Cambridge: Cambridge University Press.

Kroedel, Thomas und Schulz, Moritz (2016): „Grounding mental causation“. In: *Synthese* 193 (6), S. 1909–1923.

Laudan, Larry (1981): „A confutation of convergent realism“. In: *Philosophy of Science* 48 (1), S. 19–49.

Levin, Janet (2018): „Functionalism“. In: E. N. Zalta (Hrsg.): *The Stanford Encyclopedia of Philosophy* (Fall 2018 Edition). URL = <https://plato.stanford.edu/archives/fall2018/entries/functionalism/>.

Lewis, David (1973a): *Counterfactuals*. Oxford: Blackwell.

Lewis, David (1973b): „Causation“. In: *Journal of Philosophy* 70 (17), S. 556–567.

Lewis, David (1979): „Counterfactual Dependence and Time's Arrow“. In: *Noûs* 13 (4), S. 455–476.

Lewis, David (1983): „New Work for a Theory of Universals“. In: *Australasian Journal of Philosophy* 61 (4), S. 343–377.

Lewis, David (1986a): *On the plurality of worlds*. Oxford: Blackwell.

Lewis, David (1986b): „Postscripts to ‘causation’“. In: D. Lewis: *Philosophical Papers Vol. II*. Oxford: Oxford University Press. S. 172–213.

Lewis, David (1986c): „Events“. In: D. Lewis: *Philosophical Papers Vol. II*. Oxford: Oxford University Press. S. 241–270.

Lewis, David (1994): „Reduction of mind“. In: S. Guttenplan (Hrsg.): *Companion to the Philosophy of Mind*. Oxford: Blackwell. S. 412–431.

Lewis, David (2000): „Causation as influence“. In: *Journal of Philosophy* 97 (4), S. 182–197.

Lewis, David (2015): „Reply to Dana Scott. Is There Life on Possible Worlds?“. In: B. Loewer und J. Schaffer (Hgg.): *In Companion to David Lewis*. Oxford: Wiley-Blackwell. S. 18–22.

List, Christian und Menzies, Peter (2009): „Non-Reductive Physicalism and the Limits of the Exclusion Principle“. In: *The Journal of Philosophy* 9, S. 475–502.

List, Christian and Menzies, Peter (2017): „My brain made me do it: The exclusion argument against free will, and what's wrong with it“. In: H. Beebee, C. Hitchcock und H. Price (Hgg.): *Making a Difference*. Oxford: Oxford University Press. S. 269–285.

Loewer, Barry (2002): „Comments on Jaegwon Kim's Mind and the Physical World“. In: *Philosophy and Phenomenological Research* 65 (3), S. 655–662.

Loewer, Barry (2007): „Mental Causation, or something near enough“. In: B. P. McLaughlin und J. D. Cohen (Hgg.): *Contemporary Debates in Philosophy of Mind*. Oxford: Blackwell. S. 243–263.

Loewer, Barry (2009): „Why is there anything except physics?“. In: *Synthese* 170 (2), S. 217–233.

Loewer, Barry (2015): „Mental Causation. The free Lunch“. In: T. Horgan, M. Sabatés und D. Sosa (Hgg.): *Qualia and mental causation in a physical world : themes from the philosophy of Jaegwon Kim*. Cambridge: Cambridge University Press. S. 40–63.

Lowe, E. J. (2000): „Causal closure principles and emergentism“. In: *Philosophy* 75 (294), S. 571–586.

Lyons, Jack C. (2006): „In defense of epiphenomenalism“. In: *Philosophical Psychology* 19 (6), S. 767–794.

Macdonald, Cynthia und Macdonald, Graham (1986): „Mental Causes and the Explanation of Action“. In: *Philosophical Quarterly* 36 (143), S. 145–158.

Macdonald, Cynthia und Macdonald, Graham (2006): „The metaphysics of mental causation“. In: *Journal of Philosophy* 103 (11), S. 539–576.

Mach, Ernst (1896): *Die Principien der Wärmelehre*, Leipzig: Barth.

Mackie, John L. (1965): „Causes and Conditions“. In: *American Philosophical Quarterly* 2 (4), S. 245–264.

Maurin, Anna-Sofia (1998): „Davidson on properties“. In: *Dialectica* 52 (1), S. 13–22.

McCain, Kevin (2015): „Interventionism Defended“. In: *Logos and Episteme* 6 (1), S. 61–73.

McDermott, Michael (2007): „True antecedents“. In: *Acta Analytica* 22 (4), S. 333–335.

McDonnell, Neil (2015): „The Deviance in Deviant Causal Chains“. In: *Thought: A Journal of Philosophy* 4 (2), S.162–170.

McDonnell, Neil (2017): „Causal exclusion and the limits of proportionality“. In: *Philosophical Studies* 174 (6), S. 1459–1474.

McGrath, Sarah (2005): „Causation By Omission: A Dilemma“. In: *Philosophical Studies* 123 (1–2), S. 125–148.

McLaughlin, Brian and Bennett, Karen (2018): „Supervenience“. In: E. N. Zalta (Hrsg.): *The Stanford Encyclopedia of Philosophy* (Winter 2018 Edition), URL = <https://plato.stanford.edu/archives/win2018/entries/supervenience/>.

Mele, Alfred (2013): „Actions, Explanations, and Causes“. In: A. Laitinen, C. Sandis und G. D’Oro (Hgg.): *Reasons and Causes: Causalism and Non-causalism in the Philosophy of Action*. London: Palgrave-Macmillan. S. 160–174.

Melnyk, Andrew (2003a): *A Physicalist Manifesto: Thoroughly Modern Materialism.* Cambridge: Cambridge University Press.

Melnyk, Andrew (2003b): „Some evidence for physicalism“. In: S. Walter und H.-D. Heckmann (Hgg.): *Physicalism and Mental Causation. The Metaphysics of Mind and Action.* Exeter: Imprint Academic. S. 155–172.

Melnyk, Andrew (2016): „Grounding and the Formulation of Physicalism“. In: K. Aizawa und C. Gillett (Hgg.): *Scientific Composition and Metaphysical Ground*. London Palgrave Macmillan. S. 249–269.

Menzies, Peter (1988): „Against Causal Reductionism“. In: *Mind* 97, S. 551–574.

Menzies, Peter (2004): „Difference-making in context“. In: J. Collins, N. Hall und L. Paul (Hgg.): *Causation and Counterfactuals*. Cambridge, MA: MIT Press. S. 139–180.

Menzies, Peter (2013): „Mental causation in the physical world. In: S. C. Gibb, E.J. Lowe und R. Ingthorsson (Hgg.): *Mental Causation and Ontology*. Oxford: Oxford University Press. S. 58–87.

Menzies, Peter (2015): „The Causal Closure Argument is No Threat to Non-Reductive Physicalism“. In: *Humana.Mente. Journal of Philosophical Studies* 8 (29), S. 21–46.

Menzies, Peter und Price, Huw (1993): „Causation as a secondary quality“. In: *British Journal for the Philosophy of Science* 44 (2), S. 187–203.

Montero, Barbara (1999): „The Body Problem“. In: *Noûs* 33 (2), S. 183–200.

Montero, Barbara (2003): „Varieties of causal closure“. In: S. Walter und H.-D. Heckmann (Hgg.): *Physicalism and Mental Causation. The Metaphysics of Mind and Action.* Exeter: Imprint Academic. S. 173–187.

Moore, Dwayne (2017): „Mental causation, compatibilism and counterfactuals“. In: *Canadian Journal of Philosophy* 47 (1), S. 20–42.

Morris, Kevin (2014): „Causal Closure, Causal Exclusion, and Supervenience Physicalism“. *Pacific Philosophical Quarterly* 95 (1), S. 72–86.

Morris, Kevin (2015): „Against Disanalogy-Style Responses to the Exclusion Problem“. In: *Philosophia* 43 (2), S. 435–453.

Morris, Kevin (2018): „Physicalism, Truthmaking, and Levels of Reality: Prospects and Problems“. In: *Topoi* 37 (3), S. 473 – 482.

Morris, Kevin (2019): *Physicalism Deconstructed: Levels of Reality and the Mind–Body Problem.* Cambridge: Cambridge University Press.

Mumford, Stephen und Anjum, Rani Lill (2009): „Double prevention and powers“. In: *Journal of Critical Realism* 8 (3), S. 277–293.

Mumford, Stephen und Anjum, Rani Lill (2011): *Getting Causes From Powers.* Oxford: Oxford University Press.

Ney, Alyssa (2016a): „Grounding in the Philosophy of Mind: A Defense“. In: K. Aizawa und C. Gillett (Hgg.): *Scientific Composition and Metaphysical Ground*. London: Palgrave Macmillan. S. 271–300.

Ney, Alyssa (2016b): „Microphysical Causation and the Case for Physicalism“. In: *Analytic Philosophy* 57 (1), S. 141–164.

Ney, Alyssa (2020): „Mind“. In: M. Raven (Hrsg.): *The Routledge Handbook of Metaphysical Grounding*. New York: Routledge. S. 460–471.

Nimtz, Christian (2009): „‚Physisches' und Multi-Realisierbarkeit, oder: zwei Probleme für den Physikalismus gelöst“. In: J. Michel (Hrsg.): *Physikalismus – Willensfreiheit – Künstliche Intelligenz.* Paderborn: mentis. S. 23–42.

Oppenheim, Paul und Hilary, Putnam (1958): „Unity of Science as a Working Hypothesis“. In: *Minnesota Studies in the Philosophy of Science* 2. S. 3–36.

Papineau, David (2001): „The Rise of Physicalism“. In: C. Gillett und B. Loewer (Hgg.): *Physicalism and its Discontents.* Cambridge: Cambridge University Press. S. 3–36.

Papineau, David (2002): *Thinking About Consciousness.* Oxford: Oxford University Press.

Papineau, David (2008): „Must a physicalist be a microphysicalist?“. In: J. Hohwy und J. Kallestrup (Hgg.): *Being Reduced: New Essays on Reduction, Explanation, and Causation.* Oxford: Oxford University Press, S. 126–148.

Papineau, David (2009): „The Causal Closure of the Physical and Naturalism“. In: B. McLaughlin, A. Beckermann und S. Walter (Hgg.): *The Oxford Handbook of Philosophy of Mind*. Oxford: Oxford University Press. S. 53–65.

Pauen, Michael, Staudacher, Alexander und Walter, Sven (2006): „Epiphenomenalism: Dead end or way out?“. In: *Journal of Consciousness Studies* 13 (1–2), S. 7–19.

Paul, L.A. und Hall, Ned (2013): *Causation: A User's Guide.* Oxford: Oxford University Press.

Paull, R. Cranston und Sider, Theodore R. (1992): „In defense of global supervenience“. In: *Philosophy and Phenomenological Research* 52 (4), S. 833–854.

Pearl, Judea (2000): *Causality: Models, Reasoning, and Inference.* Cambridge: Cambridge University Press.

Penczek , Alan (1997): „Counterfactuals with true components“. In: *Erkenntnis* 46 (1), S. 79–85.

Pernu, Tuomas (2013): „Does the Interventionist Notion of Causation Deliver Us from the Fear of Epiphenomenalism?“. In: *International Studies in the Philosophy of Science* 27 (2), S. 157–172.

Pernu, Tuomas (2016): „Causal Exclusion and Downward Counterfactuals“. In: *Erkenntnis* 81 (5), S. 1031–1049.

Petrie, Bradford (1987): „Global supervenience and reduction“. In: *Philosophy and Phenomenological Research* 48, S. 119–130.

Pettit, Philip (1993): „A definition of physicalism“. In: *Analysis* 53 (4), S. 213–223.

Place, Ullin T. (1956): „Is consciousness a brain process“. In: *British Journal of Psychology* 47 (1), S. 44–50.

Platon: *Sophistes.* URL = <http://www.opera-platonis.de/Sophistes.pdf>. Abgerufen am 07.06.2020.

Poland, Jeffrey (1994): *Physicalism, the Philosophical Foundations.* Oxford: Oxford University Press.

Polger, Thomas W. (2009): „Identity theories“. In: *Philosophy Compass* 4 (5), S. 822–834.

Polger, Thomas W. (2015): „Realization and Multiple Realization, Chicken and Egg“. In: *European Journal of Philosophy* 23 (4), S. 862–877.

Polger, Thomas W. und Shapiro, Lawrence A. (2016): *The Multiple Realization Book*. Oxford: Oxford University Press.

Polger, Thomas W., Shapiro, Lawrence A., und Stern, Reuben (2018): „In defense of interventionist solutions to exclusion“. In: *Studies in History and Philosophy of Science Part A* 68, S. 51–57.

Price, Huw und Corry, Richard (Hgg.) (2007): *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited.* Oxford: Oxford University Press

Putnam, Hilary (1967): „Psychological predicates“. In: W. H. Capitan und D. D. Merrill (Hgg.): *Art, Mind, and Religion.* Pittsburgh: University of Pittsburgh Press. S. 37 – 48. [Wiederabdruck in: Heil, John (Hrsg.): *Philosophy of Mind. A guide and anthology.* Oxford: Oxford University Press 2004, S. 158–167.]

Putnam, Hilary (1975): „Philosophy and Our Mental Life“. In: Putnam, Hilary: *Mind, Language, and Reality*. Cambridge: Cambridge University Press. S. 291–303.

Quine, Willard V. O. (1960): *Word and Object*. Cambridge, MA: MIT Press.

Quine, Willard V. O. (1985): „Events and Reification“. In: E. Lepore, B. McLaughlin (Hgg.): *Actions and Events: Perspectives on the Philosophy of Davidson.* Oxford: Blackwell. S. 162 – 171.

Raatikainen, Panu (2010): „Causation, exclusion, and the special sciences“. In: *Erkenntnis* 73 (3), S. 349–363.

Raatikainen, Panu (2018): „Kim on Causation and Mental Causation“. In: *E-Logos Electronic Journal for Philosophy* 25 (2), S. 22–47.

Rabin, Gabriel (2020) „Fundamentality Physicalism“. In: *Inquiry: An Interdisciplinary Journal of Philosophy*. S. 1–40.

Raven, Michael J. (2020a): „Introduction“. In: M.J. Raven (Hgg.): *The Routledge Handbook of Metaphysical Grounding*. New York: Routledge. S. 1–14.

Raven, Michael J. (Hrsg.) (2020b): *The Routledge Handbook of Metaphysical Grounding*. New York: Routledge.

Ravenscroft, Ian (1997): „Physical Properties“. In: *Southern Journal of Philosophy* 35 (3), S. 419–431.
Robb, David (1997): „The properties of mental causation“. In: *Philosophical Quarterly* 47 (187), S. 178–94.
Robinson, William S. (2006): „Knowing epiphenomena“. In: *Journal of Consciousness Studies* 13 (1–2), S. 85–100.
Rosen, Gideon (2010): „Metaphysical Dependence: Grounding and Reduction“. In: B. Hale und A. Hoffmann (Hgg.): *Modality: Metaphysics, Logic, and Epistemology*. Oxford: Oxford University Press. S. 109–136.
Russell, Bertrand (1912): „On the Notion of Cause“. In: *Proceedings of the Aristotelian Society* 7, S. 1–26.
Russo, Andrew (2015): „Kim’s dilemma: why mental causation is not productive“. In: *Synthese* 193 (7), S. 2185–2203.
Salmon, Wesley (1971): *Statistical Explanation & Statistical Relevance*. Pittsburgh: University of Pittsburgh Press.
Salmon, Wesley (1984): *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press
Sartorio, Carolina (2006): „Disjunctive Causes“. In: *Journal of Philosophy* 103 (10), S. 521–538.
Schaffer, Jonathan (2000): „Causation by disconnection“. In: *Philosophy of Science* 67 (2), S. 285–300.
Schaffer, Jonathan (2003a): „Is there a fundamental level?“. In: *Noûs* 37 (3), S. 498–517.
Schaffer, Jonathan (2003b): „Overdetermining causes“. In: *Philosophical Studies* 114 (1–2), S. 23 – 45.
Schaffer, Jonathan (2004a): „Two Conceptions of Sparse properties“. In: *Pacific Philosophical Quarterly* 85 (1), S. 92–102.
Schaffer, Jonathan (2004b): „Causes need not be physically connected to their effects: The case for negative causation“. In: C. R. Hitchcock (Hrsg.): *Contemporary Debates in Philosophy of Science*. Oxford: Blackwell. S. 197–216.
Schaffer, Jonathan (2009): „On what grounds what“. In: D. Manley, D. J. Chalmers und R. Wasserman (Hgg.): *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford: Oxford University Press. S. 347–383.
Schneider, Susan (2012): „Non-Reductive Physicalism Cannot Appeal to Token Identity“. In: *Philosophy and Phenomenological Research* 85 (3), S. 719–728.
Schwitzgebel, Eric (2014): „The Crazyist Metaphysics of Mind“. In: *Australasian Journal of Philosophy* 92 (4), S. 665–682.
Shapiro, Lawrence A. und Sober, Elliott (2007): „Epiphenomenalism – the do’s and the don’ts’“. In: G. Wolters und P. K. Machamer (Hgg.): *Studies in Causality: Historical and Contemporary*. Pittsburgh: University of Pittsburgh Press.
Shoemaker, Sydney (1981): „Some varieties of functionalism“. In: *Philosophical Topics* 12 (1), S. 93–119.
Shoemaker, Sydney (2007): *Physical Realization*. Oxford: Oxford University Press.
Sider, Theodore (2003): „What’s So Bad About Overdetermination?“. In: *Philosophy and Phenomenological Research* 67 (3), S. 719 – 726.
Skiles, Alexander (2020): „Necessity“. In: M. J. Raven (Hrsg.): *The Routledge Handbook of Metaphysical Grounding*. New York: Routledge. S. 148–163.

Smart, John J. C. (1959): „Sensations and brain processes“. In: *Philosophical Review* 68 (April), S. 141–156.
Stalnaker, Robert (1968): „A Theory of Conditionals“. In: N. Rescher (Hrsg.): *Studies in Logical Theory (American Philosophical Quarterly Monographs 2).* Oxford: Blackwell. S. 98–112.
Stalnaker, Robert (1996): „Varieties of supervenience“. In: *Philosophical Perspectives* 10, S. 221–242.
Statham, Georgie (2018): „Woodward and variable relativity“. In: *Philosophical Studies* 175 (4), S. 885–902.
Stoljar, Daniel (2010): *Physicalism.* London/New York: Routledge.
Stoljar, Daniel (2015): „Physicalism“. In: E.N. Zalta (Hrsg): *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition), URL = <https://plato.stanford.edu/archives/win2017/entries/physicalism/>.
Strevens, Michael (2007): „Review of Woodward, Making Things Happen“. In: *Philosophy and Phenomenological Research* 74 (1), S. 233–249.
Strevens, Michael (2008): „Comments on Woodward, Making Things Happen“. In: *Philosophy and Phenomenological Research* 77 (1), S. 171–192.
Thompson, Naomi (2020): „Strict Partial Order“. In: M.J. Raven (Hrsg.): *The Routledge Handbook of Metaphysical Grounding*. New York: Routledge. S. 259–270.
van Riel, Raphael (2014a): *The Concept of Reduction.* New York u.a.: Springer.
van Riel, Raphael (2014b): „The Natures of Types and Tokens: On the Metaphysical Commitments of Non-Reductive Physicalism“. In: *Metaphysica* 15 (1), S. 219–237.
Vicente, Agustín (2006): „On the causal completeness of physics“. In: *International Studies in the Philosophy of Science* 20 (2). S. 149–171.
Vinueza, Adam (2001): „Realism and Mind Independence“. In: *Pacific Philosophical Quarterly* 82 (1), S. 51–70.
von Wright, Georg Hendrik (1971): *Explanation and Understanding*. Ithaca: Cornell University Press.
Walter, Sven (2007): „Determinables, Determinates, And Causal Relevance“. In: *Canadian Journal of Philosophy* 37 (2), S. 217–243.
Walter, Sven (2008a): „The supervenience argument, overdetermination, and causal drainage: Assessing Kim’s master argument“. In: *Philosophical Psychology* 21 (5), S. 673–696.
Walter, Sven (2008b): „Mentale Verursachung: Kausale Exklusion als Argument gegen den nicht-reduktiven Physikalismus.“ In: M. Peschl und A. Batthyany (Hgg.): *Geist als Ursache: Mentale Verursachung im interdisziplinären Diskurs.* Würzburg: Königshausen und Neumann. S. 103–119.
Walter, Sven (2009): „Epiphenomenalism and the notion of causation“. In: M. Fürst, W. Gombocz und C. Hiebaum (Hgg.), *Gehirne Und Personen.* Frankfurt a.M.: Ontos. S. 227–236.
Walters, Lee (2016): „Possible World Semantics and True-True Counterfactuals“. In: *Pacific Philosophical Quarterly* 97 (3), S. 322–346.
Walters, Lee und Williams, Robert (2013): „An Argument for Conjunction Conditionalization“. In: *Review of Symbolic Logic* 6 (4), S. 573–588.
Weslake, Brad (2011) „Exclusion Excluded“. <URL = http://bweslake.s3.amazonaws.com/research/papers/weslake_exclusion.pdf>. Abgerufen am 08.06.2020.

Weslake, Brad (2017): „Difference-Making, Closure and Exclusion". In H. Beebee, C. Hitchcock und H. Price (Hgg.), *Making a Difference*. Oxford: Oxford University Press. S. 215–231.

Wilson, Jessica (2005): „Supervenience-based formulations of physicalism". In: *Noûs* 39 (3), S. 426–459.

Wilson, Jessica (2006): „On Characterizing the Physical". In: *Philosophical Studies* 131 (1), S. 61–99.

Wilson, Jessica (2009): „Determination, realization and mental causation". In: *Philosophical Studies* 145 (1), S. 149–169.

Wilson, Jessica (2014): „No Work for a Theory of Grounding". In: *Inquiry: An Interdisciplinary Journal of Philosophy* 57 (5–6), S. 535–579.

Wilson, Jessica (2016): „Grounding-based Formulations of Physicalism". In: *Topoi* 37 (3), S. 495 – 512.

Woodward, James (2003): *Making Things Happen: A Theory of Causal Explanation.* Oxford: Oxford University Press.

Woodward, James (2007): „Causation with a human face". In: H. Price und R. Corry (Hgg.): *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*. Oxford: Oxford University Press. S. 66–105.

Woodward, James (2008a): „Mental causation and neural mechanisms". In: J. Hohwy und J. Kallestrup (Hgg.): *Being Reduced: New Essays on Reduction, Explanation, and Causation.* Oxford: Oxford University Press. S. 218–262.

Woodward, James (2008b): „Response to Strevens". In: *Philosophy and Phenomenological Research* 77 (1), S. 193–212.

Woodward, James (2015): „Interventionism and Causal Exclusion". In: *Philosophy and Phenomenological Research* 91 (2), S. 303–347.

Woodward, James (2016): „The problem of variable choice". In: *Synthese* 193 (4), S. 1047–1072.

Woodward, James (2017): „Intervening in the Exclusion Argument". In: H. Beebee, C. Hitchcock und H. Price (Hgg.), *Making a Difference*. Oxford: Oxford University Press. S. 251–268.

Yablo, Stephen (1987): „Identity, Essence, and Indiscernibility". In: *Journal of Philosophy* 84 (6), S. 293- 314.

Yablo, Stephen (1992): „Mental causation". In: *Philosophical Review* 101 (2), S. 245–280.

Yang, Eric (2013): „Eliminativism, interventionism and the Overdetermination Argument". In: *Philosophical Studies* 164 (2), S. 321–340.

Zhong, Lei (2011): „Can Counterfactuals Solve the Exclusion Problem?". In: *Philosophy and Phenomenological Research* 83 (1), S. 129–147.

Zhong, Lei (2014): „Sophisticated Exclusion and Sophisticated Causation". In: *Journal of Philosophy* 111 (7), S. 341–360.

Zhong, Lei (2015): „Why the Counterfactualist Should Still Worry About Downward Causation". In: *Erkenntnis* 80 (1), S. 159–171.